Chemometrics and stable isotope ratios of wine

Nikola Dordevic

Chemometrics and stable isotope ratios of wine

Nikola Dordevic

Chemometrics and stable isotope ratios of wine

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. Th.L.M Engelen,
volgens besluit van het college van decanen
in het openbaar te verdedigen op dinsdag 24 november 2015
om 10.30 uur precies

door

Nikola Dordevic
geboren op 17 augustus 1980
te Smederevo, Servië

Promotor:

Prof. dr. L.M.C. Buydens

Copromotoren:

Dr. R. Wehrens (WUR)
Dr. F. Camin (Fondazione Edmund Mach, Italië)

Manuscriptcommissie:

Prof. dr. F.P.J.T. Rutjes
Prof. dr. S. van Ruth (WUR)
Dr. M. Cocchi (Università degli Studi di Modena e Reggio Emilia,Italië)

This research was performed in collaboration with the Fondazione Edmund Mach, Italy.

ISBN 978-94-92380-00-5

Content

Chapter 1

1 Introduction

1.1 Wine fraud

In general, food quality is becoming more important for consumers, manufactures and regulatory institutions latterly. Various factors have contributed to the fact that consumers are more concerned with what they eat and drink, and they want to have more information about the provenance of food, the composition and the production methods. In addition to the basic prerequisites related to health and safety, another quality feature can be found in the link of a foodstuff to its geographical origin or production technology.

Consumers consider that products originated from well-defined and known origin possess higher level of quality and control in comparison to others with unknown origin, thus they are willing to spend more money on them. The production of food in a particular region or by using traditional methods is considered to add value to the product itself on the market. Therefore, the European Union (EU) plays a key role by recognizing the relationship between the product and the area of origin as quality attribute introducing appropriate regulation in this field, which aims to promote and to protect the typical food coming from different European Union countries.

Lately, there was a huge development in food production and in the control procedures related to the food safety (European Commission 2002). The major improvement is implementation of quality labels or marks from national level of protection to the European one (European Commission 1992a, 1992b and 2006). In a context of international food trade, the possibility to trace the origin and provenance of food products through the study of their composition is of great importance, and it has the goal to minimize the possible risk of imitation, counterfeiting and adulteration. This is especially important for commodities certified by quality labels, such as PDO (Protected Designation of Origin), PGI (Protected Geographical Indication), TSG (Traditional Specialty Guaranteed) and Organic Farming. PDO label covers agricultural products and foodstuffs which are produced in a given geographical area using the defined technology. PGI label covers agricultural products and foodstuffs connected to the geographical area. At least one of the stages of production takes place in the area. TSG label emphasize traditional character, in the composition or means of production. Those commodities are usually recognized as “unique” in its category.

In Italy, all regions significantly contribute to the variety of food commodities, which are characterized by an excellent quality and by the quality denominations. These marks are of great importance for these products, and they provide consumers with a guarantee of a "local-national product" and additional confirmation that the food production chain and its

quality elements have been evaluated and accepted by the authority. By promoting local products in this manner, in addition to the contribution to local economic development, it contributes to the promotion of the culture of the given community as well. Among the wide range of agricultural and food products, these laws are applied in wine sector also.

According to food legislation, wines can be with or without a designated origin (table wines). Wines with a designated origin must be produced, processed and prepared in a specific region using traditional methods. The PDO for food and DOC (Controlled designation of origin) for wines are geographical indications to protect the names of regional foods. These products must have characteristics which are essentially due to their region of production and must be produced, processed and prepared exclusively within that region (European Commission 2006, Ministry of Agriculture, Food and Forestry 2011). This is related to the fact that climatic variations influence the process of grape growth and grape quality, which influences the structure and sensorial properties of the final product. Wine obtains its economic value from its geographical origin, which can have a significant impact on the quality of the wine.

Increasing numbers of fraud cases in the food sector have recently come to light in the media (Moore et all. 2012). What they all have in common is the fact that manufacturers pass off cheap products or ingredients as expensive ones in order to maximise profit. Declarations usually conceal the fraud, misleading the consumer. Due to globalisation of the food market, the quantity of food subject to trade between different countries is increasing steadily. While consumer choice is expanding, at the same time concerns regarding the quality and origin of food on the market are also increasing. Wine is a highly prized commodity as compared to other agricultural products, and the high price depends to a large extent on geographical origin, which affects the quality of the product. Because of this, wine has been subject to different kinds of manipulation by individual producers since ancient times (Holmberg 2010).

Economic fraud related to this aspect involves mixing high quality wine with low quality wines, often originating in other geographical regions, or making complete use of wines from other regions. The addition of beet or cane sugar to grape must or wine before or during fermentation is adopted to increase the natural ethanol content and therefore the value of wine, which commands higher prices on the market. Consumers are thus misled, as the addition of sugar is not allowed, at least in Italy. Another type of fraud involves the dilution of wine with water.

It is therefore necessary to provide reliable analytical methods that are able to detect these kinds of fraud or reduce them to a minimum.
Nowadays, many of the actual authentication systems are not based on objective criteria but mostly on certifications supported by paper declarations. There is an increase in research activities with the aim of finding objective, analytical methods able to trace authenticity of food products and some of these methods are already implemented in practice.

1.2 Wine authenticity assessment

Looking at the scientific literature, we can observe a large number of scientific articles dealing with the issue of wine authenticity assessment. Wine authenticity includes checking of several features, such as authentication of origin, grape variety, wine age, production technology and so on. It can be noted that different analytical techniques and indicators have been applied to food and wine authenticity assessment (Lees, 2003, Luykx and van Ruth, 2008, Rapaneu et al. 2009, Versari et al. 2014).

1.2.1. Traceability indicators and analytical methods in food authentication studies

Because of the fact that adulteration of wine is an ongoing problem, there is the need for suitable analytical technique to gain insight into chemical composition and changes affected by different kinds of falsification. The most important analytical techniques are based on the profiling of trace elements, phenolic compounds, volatile compounds and isotope ratios using different chromatographic and spectroscopic methods (Versari et al, 2014). An overview of the recent advances of determination of food and wine authenticity is presented here, with special focus on indicators that are successfully applied in wine authentication studies.
Regarding the geographical traceability, the analytical indicators can be classified into two main categories: direct or primary indicators and indirect or secondary indicators.
Direct or primary indicators include variables that directly connect some chemical parameter of the product with the same measured parameter of their provenience area, such as trace elements and isotopic ratios of different elements. The most commonly used direct indicators in wine authentication studies are mineral elements, stable isotope ratios of bioelements and stable isotope ratios of heavy elements.

The presence and concentrations of mineral elements reflects the geochemistry of soil of the wine growing region, as well as winemaking and storage conditions, because concentrations of some trace elements such as K, Ca, Fe, Cu, and Zn (Hopfer et al. 2013) are changing due to the some processing steps (e.g. fining, ageing), while the alkaline earth metals such as Li and Rb and some rare earth elements from the group of lanthanides, are less affected by processing and time and they are useful indicators for determination of geographical origin (Rapeanu et al. 2009). The measurements of wine trace elements are usually conducted by atomic absorption spectroscopy (AAS), inductively coupled plasma atomic emission spectrometry (ICP-AES) or ICP-mass spectrometry (ICP-MS) (Versari et al. 2014, Korenovska and Suhaj 2005). Wines from two Romanian and one Moldovan wine regions were analyzed by ICP-MS and statistical data of elemental composition was used to discriminate these wines according to grape type and geographical origin. Also, the most important trace elements for their separation were detected (Geana et al. 2013). Wines from Spain and England were evaluated using the multielemental analysis by ICP-MS and discriminant analysis. In addition, this study has indicated the power which multi-element analysis can bring to the determination of the region of origin of wines (Baxter et al. 1997). Based on the same technique, optimal discrimination was achieved and the most discriminant elements were selected for Canadian wines from two major wine producing regions (Taylor et al. 2003).This technique is also useful for other food commodities such as grated hard cheeses (Camin et al. 2012) and green coffee beans traceability (Santato et al. 2012).

The stable isotope ratios of bioelements (H, C, N, O and S) of plant materials are related to type of plant and area of growing. They are in correlation with climatic factors (humidity, temperature, amount of precipitation, etc.), as well as geographical factors (distance from the sea or other evaporation source, elevation, altitude, latitude). Stable isotope ratio analysis of H, C and O is based on measurement of the ratio between the stable isotopes of H, C and O of a product or a specific component (which do not decay over time). This type of analysis can be used to detect the addition of cane and beet sugar and water in wine and to verify the origin declared on wine labels, due to the difference in the aforementioned isotopic parameters of authentic and non-authentic wines. A more detailed explanation of the method based on these parameters is in the next section. Beside the bioelements, the isotopic analysis is extended to heavy elements as well. The link between food and their provenience area can be established by the determination of some radiogenic isotopes of heavy elements, such as strontium (Sr) (Fortunato et al.

2004, García-Ruiz et al. 2007, Swoboda et al. 2008, Trincherini et al. 2014), lead (Pb) (Larcher et al. 2003, Mihaljevic et al. 2006) and neodymium (Nd) isotopic ratios. They are presented in rocks, and thus in soil and food matrices at trace level (<1%) and allow us to trace food based on the link between the soil from the provenience area and the final product.

Based on a large scale of study of a huge number of natural mineral waters from different European countries, strontium isotope ratio ($^{87}Sr/^{86}Sr$) is confirmed to be highly relevant for geographical food authentication (Voerkelius et al. 2010). Suitable methodologies based on the strontium isotopic ratio have been developed and good concurrence of isotopic range of soils and wines has been observed (Almeida and Vasconcelos 2001, Durante et al. 2013, Marchionni et al. 2013, Durante et al. 2015, Petrini et al. 2015), showing small or no variations within the samples from the same vineyard, sampled over different years. Also, there is no change in isotopic ratios of strontium during the manufacturing processes, therefore, the isotopic composition of a processed product, such as wine, remains stable. Furthermore, strontium has the property to be assimilated by living organisms where strontium actively participates in the processes of their metabolism. During the intake of plants, there is no significant isotopic fractionation of strontium present in the soil. In comparison to the isotopic composition of bioelements (H, C and O), which in addition to their usefulness for determination of geographical origin, have application for detecting other common fraud in wine sector such as addition of sugar and water during winemaking process, strontium can be used for determination of origin only.

Improved geographical differentiations between samples can be obtained combining stable isotope ratios of bioelements with strontium. Such approaches are applied for regional origin assignment of butter (Roßman et al. 2000) and cows' milk produced within Australasia (Crittenden et al. 2007). Combination of stable isotope abundance ratios of H, C, N and S with $^{87}Sr/^{86}Sr$ has been applied in order to determine and verify the geographical origin of orange juice (Rummel et al. 2010).

Indirect or secondary indicators are related to the compositional properties of the food matrices and to the production process. They are based on the characterization of the matrix which involves measurement of a large number of parameters (chemical, physical or microbiological) which help us discriminate samples from different groups. Different analytical techniques can be used for that (Peres et al. 2007). For example, using high-throughput screening spectroscopic techniques such as infrared (IR) and nuclear magnetic resonance (NMR) spectroscopy, it is possible to characterize food product of interest

measuring of large number of chemicals in our food matrix, which represent a characteristic “fingerprint” of analyzed product. Afterwards, useful information that is “hidden” among large number of other parameters needs to be extracted using the chemometrics methods. That information is used to construct models that are able to discriminate our product of interest. For example, near-infrared (NIR) and mid-infrared (MIR) spectroscopies are fast, cheap and non-destructive techniques which can be implemented for first, routine check of food products as well as for regular food authentication analysis.

Infrared radiation is the region of the electromagnetic spectrum between the visible and the microwave wavelengths. For NIR, the nominal range of wavelengths is between 750 and 2,500 nm, while for the MIR, the spectral range is from 2,500 to 25,000 nm. Good discrimination was achieved for Trappist beer using near infrared spectroscopy coupled with multivariate data analysis (Di Egidio et al. 2011, Engel et al. 2012). Application of the same technique was explored as a fast screening method for the detection of adulteration of skim milk powder (Capuano et al. 2015). The potential of NIR spectroscopy to determine the geographical origin of honey samples originated from different countries was evaluated performing various spectral pre-treatments and classification techniques (Woodcock et al. 2007, Hennessy et al 2008, and Woodcock et al. 2009). There are numerous studies of applications of near-infrared (NIR) and mid-infrared (mid-IR) in wine and other beverage industries. Applicability of mentioned techniques in combination with the multivariate data analysis was evaluated for compositional analysis, process monitoring, geographical origin, grape variety and sensory profiling (Gishen et al. 2010, Cozzolino et al. 2003, Cozzolino et al. 2011a, Cozzolino et al. 2011b, Cozzolino et al. 2011c, Cozzolino et al. 2011d, Culbert et al. 2015). They offer possibilities for cheap, simple, fast and on-line testing, including measurement of bottled product in situ.

NMR spectra are reliable source of information with a large potential in food authentication applications. The main advantage is that in a single NMR experiment a large amount of information can be obtained, measuring small molecules and macromolecules simultaneously. That is why NMR is currently recognized as an important analytical technique for food and thus in wine authentication area. Hydrogen nuclear magnetic resonance (^{1}H-NMR) spectroscopy has been applied to detect biomarkers of botanical and geographical origin for Corsican honey (Donarski et al. 2008a) and to classify honey into two groups by geographical origin (Donarski et al. 2008b). The same technique confirmed to be valuable source of information for discrimination between Ligurian and non-Ligurian

olive oils (Mannina et al. 2010). In wine authentication studies, ^{1}H-NMR spectroscopy is recognized as a very promising tool for geographical discrimination for several PDO Lambrusco red wines produced in the province of Modena, in Italy (Papotti et al. 2013). The same source of information is confirmed to be a good tool for predicting the geographical origin and shows that classification models can be improved when ^{1}H-NMR profiles are fused with stable isotope (SNIF-NMR, ^{18}O, ^{13}C) data where perfect separation is obtained for samples originating from neighboring wine regions in Germany (Monakhova et all. 2014). Discrimination between red wines obtained from organic and biodynamic production practice was achieved and an insight about the differences in composition between them was obtained by means of ^{1}H-NMR and metabonomic investigation. Vintage and vinification procedures were found to be the most important factors that affect most of the measured parameters, while vineyard management has more limited impact on wine composition (some aminoacids, alcohols and some polyphenols) (Laghi et al. 2014).

Aroma of wines can be affected by different factors, such as climate, grape variety, yeast strain, fermentation and storage conditions. The volatile compounds can be analyzed by different analytical techniques such as proton transfer reaction mass spectrometry (PTR-MS), gas chromatography mass spectrometry (GC-MS), gas chromatography (GC) and sensor or electronic nose technology.

PTR-MS enables quantitative on-line monitoring of volatile organic compounds such as alcohols, alkenes, aldehydes, ketones, nitriles, sulphides and many others) in complex matrices with detection limits as low as a few parts per trillion. High classification rate was obtained for olive oil samples originated from five different countries based on PTR-MS analysis coupled with chemometrics (Araghipour et al. 2008). By using the same analytical platform, discrimination between Dutch PDO cheese and other commercial Dutch brands without PDO protection were achieved and most important volatiles for their discrimination were detected (Galle et al. 2011). Among the successful applications of wide range of food matrices, this technique has limited application for alcoholic beverages due to the interference of ethanol. According to this technique, separation of different French wines, on the basis of the respective grape variety and region of origin, is possible only on the basis of esters after eliminating undesired effect of ethanol (Romano et al. 2014).

An automated procedure coupled with gas chromatography–ion trap mass spectrometry (GC–ITMS), was developed and employed for fast characterization of honey (Cajka et al.

2007, Cajka et al. 2009), olive oil (Cajka et al. 2010a) and beer volatiles (Cajka et al. 2010b) with the aim to confirm the authenticity of PDO commodities.

In wine origin assessment, analysis of volatile compounds (aldehydes, esters, acetals, phenols etc.) are used to obtain differentiation between samples from different regions (García-Jares and Médina 1993, Diaz et al. 2003, Fernandez de Simon et al. 2008). Furthermore, particular volatile compounds are connected to some grape varieties, leading to their specific aromatic profile (Nasi et al. 2008). The polyphenolic profile of wines also serves as a valuable source of information for wine authentication based on botanical origin (Catharino et al. 2006). There are studies which confirm usability of the amino acidic profile of wines, proving that the amino acid profile can be a useful source of information for discriminating wine samples produced in different regions of Greece and Portugal (Vasconcelos and Chaves das Neves 1989, Bouloumpasi et al. 2003).

However, caution is indispensable with choosing the parameters to be used in these applications. Storage conditions and duration have significant influence on wine quality and concentration of some compounds. For example, pigments and phenolic compounds may change drastically under the unfavorable storage conditions (Arapitsas et al. 2014).

To tackle some critical aspects of the grape variety recognition and detection of artificial flavors, molecular-genomics approach is a useful tool. Nowadays, the availability of the sequence of the grapevine genome and the continuous development of molecular descriptors for wines, combined with novel techniques for genotyping and quantification, increase the possibilities to evaluate the grape variety composition throughout the enological chain. With help of DNA and RNA based test, genetic identity of plant materials can be analyzed. Grapes are characterized with a heterozygous genome and DNA polymorphisms that can be easily detected by using polymerase chain reaction (PCR) technique. PCR is a biological technique which enables us to detect low quantities of nucleic acid probes and the determination of their sequence via the amplification of DNA or RNA individual strains. Mentioned technique is used to identify the species of origin in foods (This et al. 2004, Pereira et al. 2011). Comprehensive molecular characterisation of the grape germplasm can provide insight into genetic diversity and subdivision of grapes within the European germplasm (Emanuelli et al. 2010, Emanuelli et al. 2013).

Wine authenticity assessment involves verification of several claims, such as the declaration of origin, vintage and production technology, including checks of the addition of unpermitted substances during the production process, such as sugar and/or water, the use of an inappropriate grape variety etc. Wine is a mixture of complex chemical

compounds undergoing changes due to different storage conditions, while the wine isotopic profile is stable over time and thus represents a good source of information for authenticity purposes.

Despite of a lot of promising results obtained with the before mentioned techniques, there is no study where all wine authenticity features are evaluated simultaneously. In addition to the possibility of determining the origin of the wine and the grape varieties which the largest number of studies is dealing with, it is important to mention that the method is effective for detection of other typical frauds in wine sector, such as the detection of added water and sugar during the manufacturing process. Wines undergo a chemical change of their compounds due to inadequate storage conditions that do not change wine in terms of authenticity, although quality performance can decrease significantly. Thus, it is of ultimate importance that traceability indicators from proposed method are stable over time, and therefore useful in real context of wine authentication. To have a fair comparison of all mentioned techniques with the official one, the above mentioned facts should be considered and that is not our intention here.

Isotopic profile based techniques are implemented for the official EU and change of complete regulation is a slow process. In this thesis we investigate possibilities for improvements of official method with small modification by applying advanced statistical methods (compared with the official ones) for detecting common frauds in wine sector (false claim of origin, sugar and water addition during winemaking process).

Isotopic ratios of H, C and O are affected by climate and geographical location so database with their ranges must be updated every year which is an expensive process. Knowing those relationships is important for developing a prediction models able to predict isotopic values from climate and location, and thus save money from yearly sampling process.

1.3. Wine authenticity assessment using the stable isotope ratios of wine

Isotopes are variants of the same chemical elements, which are characterised by the same number of protons and a different number of neutrons. The term isotope has its origin in the Greek words isos (ἴσος, equal) and topos (τόπος, place), which means "the same place". Thus, most elements exist in more than one form. For example, hydrogen exists as ^{1}H and ^{2}H (deuterium); oxygen exists in three different forms, as ^{16}O, ^{17}O and

^{18}O; and carbon in two different forms, as ^{12}C and ^{13}C. Their isotopic ratios ($^{2}H/^{1}H$, $^{13}C/^{12}C$ and $^{18}O/^{16}O$) in vegetable compounds have variations caused by climatic factors and the photosynthetic cycle of the plant. Due to the evaporation and condensation process of the water cycle in nature, fractionation of oxygen and hydrogen isotopes occurs. Water evaporates from the oceans (where isotopic values are more or less constant, these values being used as a standard), lakes and rivers to the atmosphere as vapour and clouds, passing through isotope fractionation. As a consequence, the isotope composition changes due to depletion of heavier isotopes (^{2}H and ^{18}O). Plant physiology also affects isotopic fractionation, with aperture and closing of stomata, which is used to adapt to the availability of soil and atmospheric water in the plant environment.

The D/H ratio of grape sugars has given us an insight into the isotopic content of groundwater taken in by the vine from the soil and through irrigation during the sugar biosynthesis period, when the grape develops (Bigwood et al. 1998). During fermentation, the deuterium contained in the sugar is distributed in the methyl [$(D/H)_1$] and methylene [$(D/H)_2$] group of the ethanol molecule (Figure 1). The $(D/H)_1$ ratio is affected by the addition of exogenous sugar (beet or cane) to the must before and during fermentation, because they have significantly different values (Table 1). The $(D/H)_2$ ratio is less affected by this kind of fraud and more affected by fermentation water. Both $(D/H)_1$ and $(D/H)_2$ are to some extent related to the geographical origin (source water) of the plant and to the vegetable species which synthesise the sugar. Using the biochemical pathways adopted for CO_2 fixation, we can differentiate the C_3 and C_4 groups of plants. Sugar from plants with a C_3 pathway, such as vines and fruit trees, is characterised by different $^{2}H/^{1}H$ and $^{13}C/^{12}C$ values as compared to sugar from a C_4 plant such as sugar cane (O'Leary 1993). C_3 plants (wheat, vines and sugar beet) fix CO_2 directly through the enzyme ribulose-biphosphate-carboxylase (Calvin pathway) while C_4 plants (maize and sugar cane) use the enzyme phosphoenol pyruvate-carboxylase (Hatch-Slack pathway). The R parameter represents the relative distribution of deuterium in methyl and methylene molecules ($R= 2(D/H)_2 / (D/H)_1$). The aforementioned parameters are determined using Site Specific Isotope Fractionation- Nuclear Magnetic Resonance (SNIF-NMR) of deuterium from wine alcohol (European Commission 1990).

The $^{13}C/^{12}C$ ratio in plant material is mainly influenced by the plant CO_2 fixation pathway. The usefulness of the carbon isotope ratio is as an additional tool to improve detection of certain kinds of fraud, such as the addition of cane sugar (Table 1). Thus it is more

effective to use the combination of both parameters, rather than one alone in order to detect any possible type of adulteration (Roßmann et al. 1996).

The $^{18}O/^{16}O$ ratio of wine water is another parameter used in wine authenticity assessment. Indeed vegetal water is significantly enriched in ^{18}O as compared to local tap water taken in by plants and eventually used to dilute wine due to the evapo-transpiration processes occurring in plants (Rossmann et al., 1999). The water cycle evaporation and condensation process influences the fractionation of oxygen and hydrogen isotopes. The changing of water from a liquid in the ocean to a vaporous state in the atmosphere influences isotope fractionation and thus the depletion of heavier isotopes in the atmosphere. The whole process is dependent on temperature and the trajectory from the water source (sea, ocean). The result of this phenomenon is a distinctive isotope pattern for meteoric water, caused by different climatic and geographical conditions (Rozanski et al.1993). The isotope ratio of oxygen ($^{18}O/^{16}O$), influenced by meteorological parameters during grape development, is a useful parameter for detection of added water in wines (Table 1) and a useful tool for origin and vintage recognition (Hermann and Voerkelius 2008, West et al. 2007).

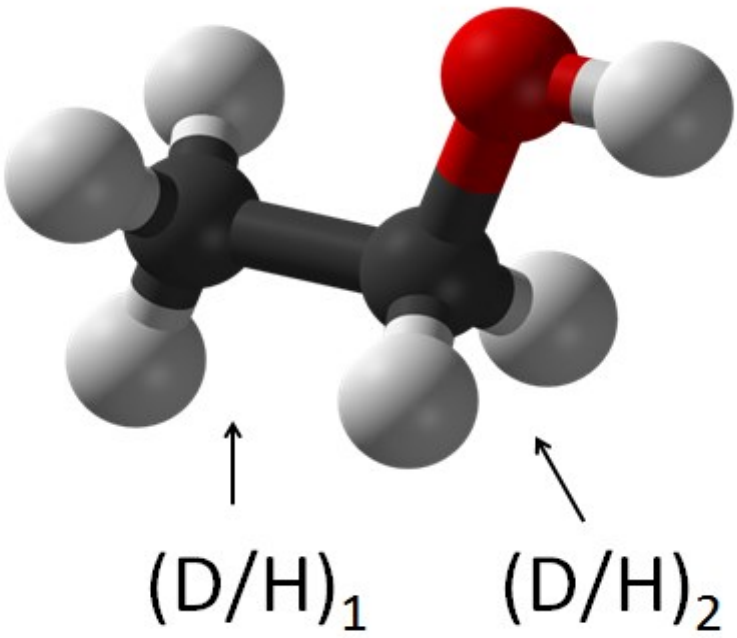

Figure1: Positions of the site specific D/H isotope ratio in the methyl and methylene position of the wine ethanol molecule, $(D/H)_1$ and $(D/H)_2$.

Table 1. Typical values for the isotopic ratios of wines (95% confidence interval) and the values for tap water, beet and cane sugar (Dordevic et al. 2013).

	$(D/H)_1$	$(D/H)_2$	$\delta^{13}C$	$\delta^{18}O$
Wines, 95% CI	98.8 to 106	124.5 to 135.5	-29.3 to -24.3	-1.3 to 8.9
Beet sugar	92.5		-27.5	
Cane sugar	109.5		-12	
Water (Italy)				-9 to -5

There are official EC/OIV methods based on stable isotope ratio analysis (Organisation Internationale de la Vigne et du Vin 2001, 2009, 2010; European Commission 2008) usable to detect fraud. In these methods the abundance of $^{13}C/^{12}C$ and $^{18}O/^{16}O$ is typically presented using delta notation ($\delta^{13}C$ and $\delta^{18}O$), with stable isotope abundance being expressed in relation to a standard.

$\delta = (R_{sample}/R_{standard} - 1) \times 1000‰$

R_{samp} = $^{13}C/^{12}C$ or $^{18}O/^{16}O$ (molar ratio of heavy and light isotopes)
The internationally recognised standards are shown in Table 2.

Table 2: Internationally recognised 13C/12C (C) and 18O/16O (O) standards

C	Pee Dee Belemnite (PDB)	R_{stand} = 0.0112372
O	Standard Mean Ocean Water (SMOW)	R_{stand} = 0.0020052

$\delta^{13}C$ and $\delta^{18}O$ parameters are measured using Isotope Ratio- Mass Spectrometry (IRMS).

1.3.1. The official EC/OIV method

In the case of must and wine, since 1991 the addition of water and exogenous sugars has been detected by analysing the isotopic ratios of hydrogen (D/H) and carbon ($^{13}C/^{12}C$) in ethanol and oxygen ($^{18}O/^{16}O$) in water. OIV (International Organization of Vine and Wine) methods are currently adopted: OIV-MA-AS311-05 for site-specific analysis of the D/H ratio using 2H-SNIF-NMR (2H-site-specific Natural Isotope Fractionation - Nuclear Magnetic Resonance), OIV-MA-AS312-07 for analysis of the $^{13}C/^{12}C$ ratio (expressed as

$\delta^{13}C$ ‰) using Isotope Ratio Mass Spectrometry (IRMS), OIV-MA-AS2-12 for analysis of the $^{18}O/^{16}O$ ratio (expressed as $\delta^{18}O$ ‰) using IRMS. Addition is detected by comparing the results with an appropriate databank, such as the official EU databank set up in 1991 (EU Regulations 2347 and 2348/91) by the European Union for all wine-producing countries within its territory. According to current Regulations (EU Regulation 555/2008), a number of samples representing the wine production of each member state are officially collected every year and analysed using the above-mentioned OIV methods. In Italy, more than 400 samples are collected every year. For each sample about 10 kilos of fresh grapes are harvested and vinified under controlled conditions. The whole process of vinification is performed without the addition of any unpermitted substances (e.g. water, beet and cane sugar), and only *Saccharomyces cerevisiae* yeast is added to complete fermentation. Finally, the resulting wines are analysed and their isotopic ratios are stored in the Wine Databank. The ultimate aim of the databank is to collect the wines' H, C and O isotope ratios ($(D/H)_1$, $(D/H)_2$, R, $\delta^{13}C$ and $\delta^{18}O$) for ethanol and wine water. The isotope data bank makes reference data available on a yearly basis, thus allowing legal limits to be defined on the basis of isotopic data for each country, each sub-area (e.g. region) and each protected denomination (PDO-IGP), as well as general limits, when origin and year of production are not declared.

1.3.2. Background

In the last 20 years many research groups have investigated parameters correlated with the isotopic values of wine.

Significant differences in $(D/H)_1$ were found in samples collected from 1987 to 1992 from the Adriatic and Tyrrhenian coasts, demonstrating the significance of the longitude factor (Monetti et al. 1994). Taking into consideration samples from small areas of Italy, positive correlation between the $(D/H)_1$ of ethanol with mean temperature and negative correlation between the same parameter and amount of precipitation were observed (Aghemo et al. 2011). Based on analysis of 284 wine samples originating in 19 different countries, correlation between $(D/H)_1$, $(D/H)_2$ and latitude were found, except in samples where irrigation took place with water coming from other regions (Bigwood et al. 1998).

Furthermore, according to work based on samples from 34 regions across the world, $(D/H)_1$ and $(D/H)_2$ and the $\delta^{13}C$ of ethanol, there is a positive correlation with mean temperature, and a negative correlation with the amount of precipitation in the period of sugar accumulation and grape ripening (Martin and Martin 2003). Based on 102 samples

from Slovenia, distance from the sea was shown to be a factor influencing variations in $(D/H)_2$, $\delta^{13}C$ and $\delta^{18}O$ values (Ogrinc et al. 2001), which may be due to different climatic conditions in coastal and inland areas (Kosir et al. 2001). The isotopic ratios of H and O are affected by climatic conditions from the start of grape maturation period until harvest, more precisely by temperature and relative humidity, which control the degree of evapotranspiration. They tend to enrich the grape water with ^{2}H and ^{18}O (Bricout 1978). Another work claimed that geographical and climatic variability of the stable isotope ratios of wines is due to fact that the H and O isotopic values of grape juices are also affected by source water, which changes with latitude and distance from the sea (Clark and Fritz 1997). The $\delta^{18}O$ values of source (ground) water are analogous to the $\delta^{18}O$ values of precipitation, and regional mean values for the $\delta^{18}O$ of precipitation, taking into account all locations in our data set, are shown in Figure 2.

Precipitation during late grape development affects $\delta^{18}O$ rapidly, due to stable isotope exchange with water vapour (Ingraham and Caldwell 1999). Furthermore, there is linear correlation between the $\delta^{18}O$ of wine water and the amount of precipitation (Martin and Martin 2003). The strong climatic dependence of $\delta^{18}O$ was confirmed for around 50 samples from the western coast of the United States using multiple regressions to build a prediction model. For this purpose, the mean daily maximum temperatures and dew point temperatures in the grape ripening period (September and October) and the $\delta^{18}O$ values of local precipitation were selected as predictors (West et all. 2007). Another model was obtained based on 775 wine samples from Germany, collected over 6 different years, showing a strong relationship between wine water $\delta^{18}O$ and relative humidity for a period of 30 days before harvest (Hermann and Voerkelius 2008). Along with climatic conditions, irrigation can decrease wine $\delta^{18}O$ and $\delta^{13}C$ values, especially in warm areas (Gomez-Alonso and Garcia-Romero 2010). A possible reason is that water used for irrigation is transported from distant areas, which have different isotopic profiles as compared to the wine-growing region. Based on analysis of 80 wine samples from southern Brazil, a trend for decreasing $\delta^{18}O$ with latitude and altitude was found. Furthermore, $\delta^{18}O$ was confirmed to have a strong discriminant capacity for samples from different areas, as compared to $\delta^{13}C$ (Adami et al. 2010).

Some authors have investigated the influence of grape varieties on the stable isotope ratios of wine (especially $\delta^{18}O$ and $\delta^{13}C$), as a consequence of differences in the date of harvest and growing site, thus in the climatic conditions of the regions of provenance (Roßmann et al. 1999, Magdas et al. 2012, West et al. 2007, Adami et al. 2010). Only

Gomez-Alonso and Garcia-Romero 2010, in a local study found that $\delta^{18}O$ and $\delta^{13}C$ were affected by variety, regardless of the sampling data. Wines from early harvesting periods had higher $\delta^{18}O$ values as compared to wines from late harvesting periods (Roßmann et al. 1999), due to the difference in climatic conditions.

The carbon isotopic ratio ($^{13}C/^{12}C$) of grape sugar and thus ethanol is affected by stomatal conductance and intercellular and ambient CO_2 concentration, which is controlled by climatic factors (temperature, relative humidity, water stress, amount of precipitation) (O'Leary 1995, Scheidegger et al. 2000, Scartazza et al. 2004).

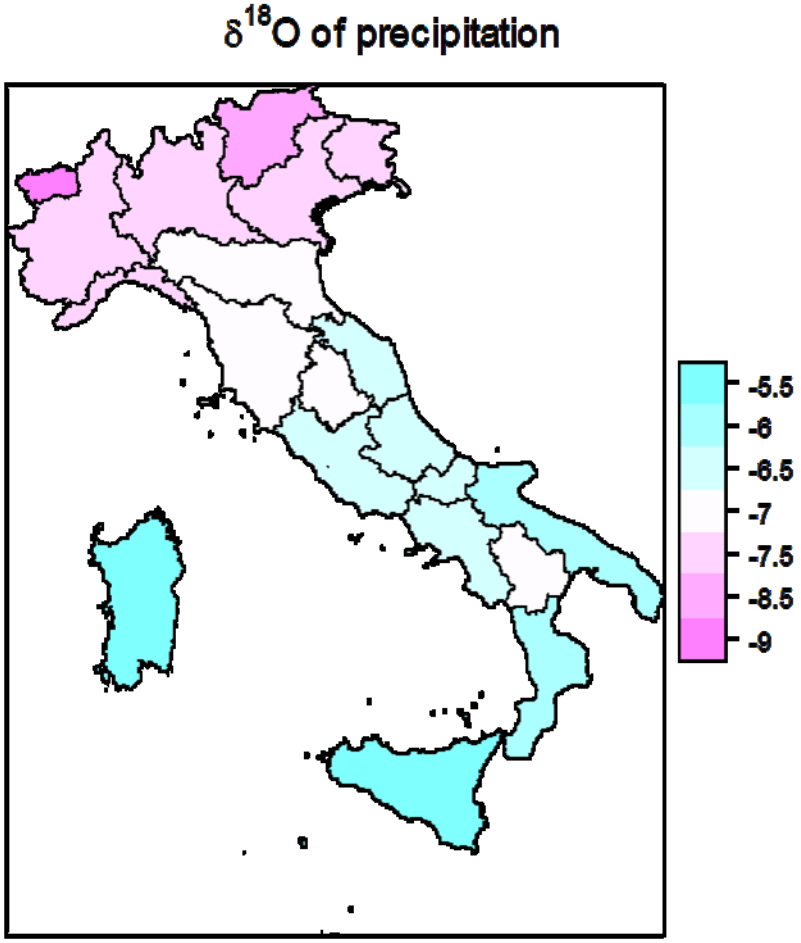

Figure 2. Regional mean values for the $\delta^{18}O$ of precipitation, taking into account all the locations in our dataset

Because of the fact that the isotope ratios of some elements vary, as a result of ecological and chemical processes caused by spatial and temporal variations of isotopes in environmental materials, these values can be predicted by models of isotope fractioning processes and data about the environmental conditions across space and time, which is called isoscape (Bowen, 2010). Isoscapes of strontium, hydrogen, oxygen, carbon and

nitrogen have been used in wide range of applications (food authentication, animal migration, climate, archeology, forensic studies, etc.) regarding the sources and origin of different matrices via their isotopic signatures. The existence of current isoscapes enables us to calculate isotopic abundances in soil, water (Bowen et al. 2005) and numerous food products, including olive oil, mineral water, honey, chicken, wheat and lamb (TRACE TOLL 2009).

1.4. Chemometrics and verification of wine authenticity claims

Chemometrics is the discipline within chemistry that develops methods to obtain relevant information from chemical data using mathematical and statistical tools. These range from data exploration to methods applied to solve predictive problems in experimental chemistry. In predictive applications, the properties of chemical systems are modelled with the intent of predicting new properties or behaviour of interest. In food authentication applications, it is important to know which group the samples belong to and the hypothesis is that the characteristics of the particular food products are described by the multivariate data structure of that group. The main task for chemometrics here is to summarise this multivariate structure appropriately in order to establish rules for correctly assessing new observations where the group membership is under evaluation.

The number of works in the field of food authentication and traceability, based on stable isotope ratios and chemometrics, is increasing. In those studies, different chemometrics methods are applied to tackle the problem of authenticity evaluation. For example, separation between olive oil samples originating in different Mediterranean countries based on $\delta^{13}C$ and $\delta^{18}O$ values was obtained by applying hierarchical cluster analysis and principal component analysis (Angerosa et al. 1999). Using multi-element and (H, C, N, S) stable isotope ratio in combination with linear discriminate analysis, a high percentage of correct classification was obtained for lamb meat from several European regions (Camin et al. 2007). With the help of canonical discriminant analysis it was confirmed that carbon and sulphur isotopic ratios provide the maximum discrimination between honey samples originating in different regions in Europe by comparing nitrogen and hydrogen isotopic ratios (Schellenberg et all. 2010). Combination of stable isotope ratios and mineral composition in tomato and derivatives along the production chain (juice, passata and

paste) were confirmed to be an adequate source for sample discrimination according to their origin by applying linear discriminant analysis (Bontempo et all. 2011).

In the field of wine authentication there are also works based on chemometrics and stable isotope ratios. The stable isotope data of wines produced in Slovenia were evaluated using principal component analysis and linear discriminant analysis. It was found that it is possible to separate wine samples originating in different regions based on $(D/H)_2$, $\delta^{13}C$, $\delta^{18}O$ vine parameters (Ogrinic et al. 2001). By using a combination of isotopic, classical parameters and elemental composition, Switzerland was divided into four main production zones with a discrimination ability that had not previously been achieved (Gremaud et al. 2004). There is only one work applying multivariate statistics to wine authentication using stable isotope ratios, based on a small set of samples from the EU Wine Databank (Wachter et al. 2009).

Most of the research papers dealing with authenticity evaluation of foods use classification techniques. These techniques are useful from scientific point of view when we want to evaluate possibilities for discrimination of two or more groups and to evaluate importance of parameters for discrimination. For the cases where it is known which parameters are representative for groups of interest, classification techniques are not very useful and class modeling techniques are more suitable. This is especially the case for the food authenticity check for PDO foods, where class modeling techniques are used to verify declared authenticity claims (labeled origin, presence of unpermitted supplements, etc.) (Forina et al. 2008, Forina et al. 2009, Casale et al. 2008). Class modelling techniques answer the question: “does a sample X, declared as belonging to class A, really belong to class A?”. This approach is different compared with the classification ones, in terms of each class being independent from the others, and the corresponding classification is more flexible. As a result, all samples under the evaluation can be accepted by the models in only one class, in more than one class or to be rejected by all class models. In univariate statistics, the answer to the class modeling technique general question can be obtained from the known significance test with the null hypothesis:

H_O: x, the value of a parameter for sample X is not significantly different from m, the estimated mean of the parameter for the class A.

While in univariate methods evaluates all parameter separately from each other, multivariate methods consider all parameters simultaneously and thus often gain a new and higher insight and quality in data evaluation. Univariate evaluation of each parameter separately would lead to the wrong conclusion that the variables are useless, while in

multivariate analysis correlations between all parameters are considered and test is more powerful (Saccenti et al. 2014).

To evaluate class model performance, the results are expressed with two parameters, sensitivity and specificity. Sensitivity value is percentage of samples from the same class which are correctly accepted by the class model (the percentage of Correct I decisions - true positives). Specificity value is the percentage of samples from all the other classes which are correctly rejected by the class model (the percentage of Correct II decisions - true negatives) (Forina et al. 2008).

For example, practical meaning is that if one red wine is labeled with particular region of origin and vintage, classification approach would require at least two groups of representative samples: the first group of wine samples should represent all authentic red wines from labeled origin and vintage, and the second group should represent all samples outside of labeled group and vintage which is not possible. Also, beside from the false claim of origin in wine sector, there are other common frauds such as addition of sugar and water during the winemaking process, etc. These fraud samples should also be included in second group which is rarely realizable. Because of these reasons, class modeling techniques are a suitable approach for addressing the majority of real authentication problems (Oliveri and Downey, 2012).

1.5 Aim and outline of the thesis

As mentioned above, the method for wine authenticity evaluation based on stable isotope ratio analysis has been in use for many years, so a large amount of data have been accumulated to date. We believe that it is time to carry out critical evaluation of the statistical methodology used to interpret data from the Wine Databank, to assess performance and possibilities for improvement, and to evaluate the climatic and geographical dependence of stable isotopes ratios used in wine authenticity evaluation. Besides extending general scientific knowledge on the stable isotope ratios of wines, the output of this thesis will provide insight into the performance of the method used in wine authenticity applications and will develop new methods for data interpretation. Furthermore, this work will help us to understand the disadvantages of the method in order to improve it in the future.

The Italian Databank is one of the best possible sources for this research because it contains a very large number of wine samples and also because these samples originate in a climatically varied country.

In the first study we investigate univariate approaches underlying the official method for each parameter separately, to assess claims of origin based on individual t-tests. To better discriminate between samples from different regions, several extensions of univariate t-tests are proposed, with one univariate and two multivariate statistical methods. The results are based on 5220 authentic Italian wine samples, collected in the period from 2000 to 2010, following the official procedure.

In the second study, the ability to detect unpermitted addition of cane and beet sugar to grapes before fermentation and the dilution of wine with water are evaluated using the univariate statistical method (based on t-tests) and its multivariate extension (Hotelling t-test). The analysis is based on validated simulations, adding different amounts of water, cane and beet sugar to wine, which provides additional insight into the importance of different isotopic parameters for fraud detection.

In the third study, the relationships between wine $(D/H)_1$, $(D/H)_2$, $\delta^{13}C$ and $\delta^{18}O$ and the climatic and geographical parameters of the provenance areas are evaluated, as well as their general climatic dependence on wine growing areas. Comprehending these relationships is important for general understanding of isotopic phenomena, which can help us to make additional improvements to wine authentication procedures based on stable isotope ratios.

Finally, chapter 5 summarises all the findings resulting from this thesis. Perspectives for further research are also presented.

References

Adami, L., Dutra, S.V., Marcon, Â.R., Carnieli, G.J., Roani, C.A. and Vanderlinde, R. (2010) Geographic origin of southern Brazilian wines by carbon and oxygen isotope analyses. Rapid Communication in Mass Spectrometry 24, 2943–2948.

Aghemo, C., Albertino, A., Gobettoa, R. and Spanna, F. (2011) Correlation between isotopic and meteorological parameters in Italian wines: a local-scale approach. Journal of the Science of Food and Agriculture 1, 2088–2094.

Almeida, C.R.C. and Vasconcelos, M.T.S.D. (2001) ICP–MS determination of strontium isotope ratio in wine in order to be used as a fingerprint of its regional origin. Journal of Analytical Atomic Spectrometry 16, 607–611.

Angerosa, F., Breas, O., Contento, S., Guillou, C., Reniero, F. and Sada, E. (1999) Application of Stable Isotope Ratio Analysis to the Characterization of the Geographical Origin of Olive Oils. Journal of Agricultural and Food Chemistry 47, 1013-1017.

Arapitsas, P., Speri, G., Angeli, A., Perenzoni, D. and Mattivi, F. (2014) The influence of storage on the ‘‘chemical age’’ of red wines. Metabolomics 10, 816–832.

Araghipour, N., Colineau, J., Koot, A., Akkermans, W., Moreno Rojas, J.M., Beauchamp, J., Wisthaler, A., Mark, T.D., Downey, G., Guillou, C., Mannina, L. and van Ruth, S. (2008) Geographical Origin Classification of Olive Oils by PTR-MS. Food Chemistry 108, 374–383.

Baxter, M.J., Crews, H.M., Dennis, M.J., Goodall, I. and Anderson, D. (1997) The determination of the authenticity of wine from its trace element composition. Food Chemistry 60, 443–450.

Bigwood, T., Sharman, M., Aldus, A. and Dennis, M.J. (1998) Global variations in the deuterium/hydrogen isotope ratios of wine. Journal of Wine Research 9:3, 155-166.

Bontempo, L., Camin, F., Manzocco, L., Nicolini, G., Wehrens, R., Ziller, L. and Larcher, R. (2011) Traceability along the production chain of Italian tomato products on the basis of stable isotopes and mineral composition. Rapid Communication in Mass Spectrometry 25, 899–909.

Bouloumpasi, E., Soufleros, E.H., Tsarchopoulos, C. and Biliaderis, C.G. (2002) Primary amino acid composition and its use in discrimination of Greek red wines with regard to variety and cultivation region. Vitis 41 (4), 195–202.

Bowen, G.J., Wassenaar, L.I. and Hobson, K.A. (2005) Global application of stable hydrogen and oxygen isotopes to wildlife forensics. Oecologia 143, 337-348.

Bowen, G.J. (2010) Isoscapes: Spatial Pattern in Isotopic Biogeochemistry. Annual Review of Earth and Planetary Sciences 38, 161-187.

Bricout, J. (1978) Recherches sur le fractionnement des isotopes stables del'hydrogene et l'oxygene dans quelques vegetaux. Rev Cytol Biol veget-Bot 1, 133-209.

Cajka, T., Hajslova, J., Cochran, J., Holadova, K. and Klimankova, E. (2007) Solid Phase Microextraction-Comprehensive Two-Dimensional Gas Chromatography-Time-of-Flight Mass Spectrometry for the Analysis of Honey Volatiles. Journal of Separation Science 30, 534–546.

Cajka, T., Hajslova, J., Pudil, F. and Riddellova, K. (2009) Traceability of Honey Origin Based on Volatiles Pattern Processing by Artificial Neural Networks. Journal of Chromatography A 1216, 1458– 1462.

Cajka, T., Riddellova, K., Klimankova, E., Cerna, M., Pudil, F. and Hajslova, J. (2010a) Traceability of Olive Oil Based on Volatiles Pattern and Multivariate Analysis. Food Chemistry 121, 282–289,

Cajka, T., Riddellova, K., Tomaniova, M. and Hajslova, J. (2010b) Recognition of beer brand based on multivariate analysis of volatile fingerprint. Journal of Chromatography A 1217, 4195–4203.

Camin, F., Bontempo, L., Heinrich, K., Horacek, M., Kelly, S.D., Schlicht, C., Thomas, F., Monahan, F,J., Hoogewerff, J. and Roßmann, A. (2007) Multi-element (H, C, N, S) stable isotope characteristics of lamb meat from different European regions. Analytical and Bioanalytical Chemistry 389, 309–320.

Camin, F., Wehrens, R., Bontempo, L., Bertoldi, D., Ziller, L., Perini, M., Nicolini, G., Nocetti, M. and Larcher R. (2012) Can H, C, N and S stable isotopes and mineral profiles objectively guarantee the authenticity of grated hard cheeses? Analytica Chimica Acta 711, 54-59.

Capuano, E., Boerrigter-Eenling, R., Koot, A. and van Ruth, S.M. (2015) Targeted and Untargeted Detection of Skim Milk Powder Adulteration by Near-Infrared Spectroscopy. Food Analytical Methods, DOI 10.1007/s12161-015-0100-3

Casale, M., Casolino, C., Ferrari, G. and Forina, M. (2008) Near infrared spectroscopy and class modelling techniques for the geographical authentication of Ligurian extra virgin olive oil. Journal of Near Infrared Spectroscopy 16, 9-4.

Catharino, R.R., Cunha, I.B., Fogaça, A.O., Facco, E.M., Godoy, H.T., Daudt, C.E., Eberlin, M.N. and Sawaya, A.C. (2006) Characterization of must and wine of six varieties of grapes by direct infusion electrospray ionization mass spectrometry. Journal of Mass Spectrometry (41) 2, 185 – 190.

Crittenden, R.G., Andrew, A.S., LeFournour, M., Young, M.D., Middleton, H. and Stockmann R. (2007) Determining the geographic origin of milk in Australasia using multi-element stable isotope ratio analysis. International Dairy Journal 17, 421–428.

Clark, I. and Fritz, P. (1997) Tracing the hydrological cycle. In: Environmental Isotopes in Hydrogeology (Lewis Publishers, Boca Raton, Florida), pp 35-78.

Culbert, J., Cozzolino, D., Ristic, R. and Wilkinson, K. (2015) Classification of Sparkling Wine Style and Quality by MIR Spectroscopy. Molecules 20, 8341-8356.

Cozzolino, D., Smyth, H.E. and Gishen, M. (2003) Feasibility Study on the Use of Visible and Near-Infrared Spectroscopy Together with Chemometrics To Discriminate between Commercial White Wines of Different Varietal Origins. Journal of Agricultural and Food Chemistry 51, 7703–7708.

Cozzolino, D., Cynkar, W.U., Shah, N. and Smith, P. (2011a) Technical solutions for analysis of grape juice, must, and wine: the role of infrared spectroscopy and chemometrics. Analytical and Bioanalytical Chemistry 401, 1475–1484.

Cozzolino, D., Cynkar, W.U., Shah, N. and Smith, P. (2011b) Multivariate data analysis applied to spectroscopy: Potential application to juice and fruit quality. Food Research International 44, 1888–1896.

Cozzolino, D., Cynkar, W.U., Shah, N. and Smith, P. (2011c) Feasibility study on the use of attenuated total reflectance mid-infrared for analysis of compositional parameters in wine. Food Research International 44, 181–186.

Cozzolino, D., Cynkar, W.U., Shah, N. and Smith, P.A. (2011d) Can spectroscopy geographically classify Sauvignon Blanc wines from Australia and New Zealand? Food Chemistry 126, 673–678.

Díaz, C., Elías Conde, J., Méndez, J. and Pérez Trujillo, J.P. (2003) Volatile compounds of bottled wines with Denomination of Origin from the Canary Islands (Spain). Food Chemistry 81, 447–452.

Di Egidio, V., Oliveri, P., Woodcock, T. and Downey, G. (2011) Confirmation of brand identity in foods by near infrared transflectance spectroscopy using classification and class-modelling chemometric techniques - The example of a Belgian beer. Food Research International 44, 544–549.

Dordevic, N., Camin, F., Marianella, R.M., Postma, G.J., Buydens, L.M.C. and Wehrens, R. (2013), Detecting the addition of sugar and water to wine. Australian Journal of Grape and Wine Research 19, 324–330.

Donarski, J.A., Jones, S.A., Harrison, M., Driffield, M. and Charlton, A.J. (2008a) Identification of Botanical Biomarkers Found in Corsican Honey. Food Chemistry 118, 987-994.

Donarski, J.A., Jones, S.A. and Charlton, A.J. (2008b) Application of Cryoprobe 1H Nuclear Magnetic Resonance Spectroscopy and Multivariate Analysis for the Verification of Corsican Honey. Journal of Agricultural and Food Chemistry 56, 5451-5456.

Durante, C., Baschieri, C., Bertacchini, L., Cocchi, M., Sighinolfi, S., Silvestri, M. and Marchetti, A. (2013) Geographical traceability based on 87Sr/86Sr indicator: A first approach for PDO Lambrusco wines from Modena. Food Chemistry 141, 2779–2787.

Durante, C., Baschieri, C., Bertacchin,L., Bertelli, D., Cocchi, M., Marchetti, A., Manzini, D., Papotti, G. and Sighinolfi, S. (2015) An analytical approach to Sr isotope ratio determination in Lambrusco wines for geographical traceability purposes. Food Chemistry 173, 557–563.

Emanuelli, F., Battilana, J., Costantini, L, Le Cunff, L., Boursiquot, J.M., This, P., Grando, M.S. (2010) A candidate gene association study on muscat flavor in grapevine (Vitis vinifera L.). BMC Plant Biology 10:241.

Emanuelli, F., Lorenzi, S., Grzeskowiak, L., Catalano, V., Stefanini, M., Troggio, M., Myles, S., Martinez-Zapater, J.M., Zyprian, E., Moreira, F.M. and Grando, M.S. (2013) Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape. BMC Plant Biology (13) 39, 1-17.

Engel, J., Blanchet, L., Buydens, L.M. and Downey, G. (2012) Confirmation of brand identity of a Trappist beer by mid-infrared spectroscopy coupled with multivariate data analysis. Talanta 99, 426-432.

European Commission (1990) Commission regulation (EEC) No 2676/90 of 17 September 1990 determining Community methods for the analysis of wines. Official Journal of the European Union L 272, 1–192.

European Commission (1991a) Commission Regulation (EEC) No 2347/91 of 29 July 1991 on the collection of samples of wine products for the purposes of cooperation between Member States and for analysis by nuclear magnetic resonance, including analysis for the purposes of the Community databank. Official Journal of the European Union L 214, 32–38.

European Commission (1991b) Commission Regulation (EEC) No 2348/91 of 29 July 1991 establishing a databank for the results of analyses of wine products by nuclear magnetic resonance of deuterium. Official Journal of the European Union L 214, 39–43.

European Commission (1992a) Council Regulation (EEC) No 2081/92 of 14 July 1992 on the protection of geographical indications and designations of origin for agricultural products and foodstuffs. Official Journal of the European Union L 208, 1–8.

European Commission (1992b) Council Regulation (EEC) No 2082/92 of 14 July 1992 on certificates of specific character for agricultural products and foodstuffs. Official Journal of the European Union L 208, 9–14.

European Commission (2002) Commission Regulation No. 178/2002 of 28 January 2002 laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety. Official Journal of the European Union L 31, 1–24.

European Commission (2006) Commission Regulation (EC) No. 510/06 on the Protection of Geographical Indications and Designations of Origin for Agricultural Products and Foodstuffs. Official Journal of the European Union L 93/12, 1-14.

European Commission (2008) Commission Regulation No. 555/2008 of 27 June 2008 laying down detailed rules for implementing Council Regulation (EC) No 479/2008 on the

common organisation of the market in wine as regards support programmes, trade with third countries, production potential and on controls in the wine sector. Official Journal of the European Union L 170, 1–80.

Fernandez de Simon, B., Cadihia, E., Sanz, M., Poveda, P., Perez-Magarino, S., Ortega-Heras, M. and Gonzalez-Huerta, C. (2008) Volatile Compounds and Sensorial Characterization of Wines from Four Spanish Denominations of Origin, Aged in Spanish Rebollo (Quercus pyrenaica Willd.) Oak Wood Barrels. Journal of Agricultural and Food Chemistry 56, 9046–9055.

Fortunato, G., Mumic, K., Wunderli, S., Pillonel, L., Bosset, J. O. and Gremaud, G. (2004) Application of strontium isotope abundance ratios measured by for authentication. Journal Of Analytical Atomic Spectrometry 19, 227-234.

Forina, M., Oliveri, P., Lanteri, S. and Casale, M. (2008) Class-modeling techniques, classic and new, for old and new problems. Chemometrics and Intelligent Laboratory Systems 93, 132–148.

Forina, M., Oliveri, P., Jäger, H., Römisch, U. and Smeyers-Verbeke, J. (2009) Class modeling techniques in the control of the geographical origin of wines. Chemometrics and Intelligent Laboratory Systems 99, 127–137.

Galle, S.A., Koot, A., Soukoulis, C., Cappellin, L., Biasioli, F., Alewijn, M. and van Ruth, S.M. (2011) Typicality and geographical origin markers of protected origin cheese from The Netherlands revealed by PTR-MS. Journal of Agricultural and Food Chemistry 59, 2554-2563.

García-Jares, C. and Médina, B. (1993) Research on white and red wine blending in the production of rosé wines by means of the partial least squares method. Journal of the Science of Food and Agriculture 63, 349–354.

García-Ruiz, S., Moldovan, M., Fortunato, G., Wunderl, S. and García Alonso, J.I. (2007) Evaluation of strontium isotope abundance ratios in combination with multi-elemental

analysis as a possible tool to study the geographical origin of ciders. Analytica Chimica Acta 590, 55-66.

Geana, I. , Iordache, A., Ionete, R. Marinescu, A., Ranca, A. and Culea, M. (2013) Geographical origin identification of Romanian wines by ICP-MS elemental analysis. Food Chemistry 138, 1125–1134.

Gishen, M., Cozzolino, D. and Dambergs, R. G. (2010) The Analysis of Grapes, Wine, and Other Alcoholic Beverages by Infrared Spectroscopy. In: Chalmers, J.M. and Griffiths P.R. (eds) Handbook of Vibrational Spectroscopy, John Wiley and Sons, Inc.

Gómez-Alonso, S. and García-Romero, E. (2010) Effect of irrigation and variety on oxygen ($\delta^{18}O$) and carbon ($\delta^{13}C$) stable isotope composition of grapes cultivated in a warm climate. Australian Journal of Grape Wine Research 16 (2), 283-289.

Gremaud, G., Quaile, S., Piantini, U., Pfammatter and E., Corvi, C. (2004) Characterization of Swiss vineyards using isotopic data in combination with trace elements and classical parameters. European Food Research and Technology 219, 97–104.

Hennessy, S., Downey, G. and O'Donnell, C. (2008) Multivariate Analysis of Attenuated Total Reflection-Fourier Transform Infrared Spectroscopic Data to Confirm the Origin of Honeys. Applied Spectroscopy 62, 1115-1123.

Hermann, A. and Voerkelius, S. (2008) Meteorological Impact on Oxygen Isotope Ratios of German Wines. American Journal of Enology and Viticulture 59 (2), 194-199.

Holmberg, L. (2010) Wine fraud. International Journal of Wine Research 2, 105–113.

Ingraham, N. and Caldwell, E. (1999) Influence of weather on the stable isotopic ratios of wines: Tools for weather/climate reconstruction. Journal of Geophysical Research 104, 2185–2194.

Hopfer, H., Nelson, J., Mitchell, A.E., Heymann, H. and Ebeler, S.E. (2013) Profiling the trace metal composition of wine as a function of storage temperature and packaging type. Journal of Analytical Atomic Spectrometry 28, 1288–1291.

Korenovska, M. and Suhaj, M. (2005) Identification of some Slovakian and European wines origin by the use of factor analysis of elemental data. European Food Research and Technology 221, 550–558.

Košir, I.J., Kocjančič, M., Ogrinc, N. and Kidrič, J. (2001) Use of SNIF-NMR and IRMS in combination with chemometric methods for the determination of chaptalization and geographical origin of wines (the example of Slovenian wines). Analytica Chimica Acta 429, 197-208.

Laghi, L., Versari, A., Marcolini, E. and Parpinello, G.P. (2014) Metabonomic Investigation by 1H-NMR to Discriminate between Red Wines from Organic and Biodynamic Grapes. Food and Nutrition Sciences 5, 52-59.

Larcher, R., Nicolini, G. and Pangrazzi, P. (2003) Isotope Ratios of Lead in Italian Wines by Inductively Coupled Plasma Mass Spectrometry. Journal of Agricultural and Food Chemistry *51* (20), 5956–5961.

Lees, M. (Editor), Food Authenticity and Traceability, Woodhead Publishing Ltd., Cambridge, UK, 2003.

Luykx, D.M.A.M. and van Ruth, S.M. (2008) An overview of analytical methods for determining the geographical origin of food products. Food Chemistry 107, 897–911.

Magdas, D.A., Cuna, S., Cristea, G., Ionete, R.E. and Costinel, D. (2012) Stable isotopes determination in some Romanian wines. Isotopes in Environmental and Health Studies 48 (2), 345-353.

Mannina, L., Marini, F., Gobbino, M., Sobolev, A.P. and Capitani, D. (2010) NMR and Chemometrics in Tracing European Olive Oils: the Case Study of Ligurian Samples. Talanta 80, 2141–2148.

Martin, G.J. and Martin, M.L. (2003) Climatic significance of isotope ratios. Phytochemistry Rev. 2, 179-190.

Marchionni, S., Braschi, E., Tommasini, S., Bollati, A., Cifelli, F., Mulinacci, N., Mattei, M. and Conticelli, S. (2013) High-Precision $^{87}Sr/^{86}Sr$ Analyses in Wines and Their Use as a Geological Fingerprint for Tracing Geographic Provenance. Journal of Agricultural and Food Chemistry 61 (28), 6822–6831.

Mihaljevic, M., Ettler, V., Sebek, O., Strnad, L. and Chrastny, V. (2006) Lead isotopic signatures of wine and vineyard soils-tracers of lead origin. Journal of Geochemical Exploration 88, 130–133.

Ministero delle politiche agricole alimentari e forestali, Dipartimento delle politiche competitive del mondo rurale e della qualita. Direzione Generale dello sviluppo agroalimentare e della qualita. SAQ IX. NUMERO DEI VINI A D.O.C.G. - D.O.C. - I.G.T. AL 19 OTTOBRE, 2011.

Monakhova, Y.B., Godelmann, R., Hermann, A., Kuballa, T., Cannet, C., Schäfer, H., Spraul, M. and Rutledge, D.N. (2014) Synergistic effect of the simultaneous chemometric analysis of 1H NMR spectroscopic and stable isotope (SNIF-NMR, 18O, 13C) data: Application to wine analysis. Analytica Chimica Acta 833, 29–39.

Moore, J.C., Spink, J. and Lipp, M. (2012) Development and Application of a Database of Food Ingredient Fraud and Economically Motivated Adulteration from 1980 to 2010. Journal of Food Science 77, R118–R126.

Nasi, A., Ferranti P., Amato, S. and Chianese, L. (2008) Identification of free and bound volatile compounds as typicalness and authenticity markers of non-aromatic grapes and wines through a combined use of mass spectrometric techniques. Food Chemistry 110, 762–768.

O'Leary M.H. (1995) Environmental Effects on Carbon Isotope Fractionation in Terrestrial Plant. In: Wada, E., Yoneyama, T., Mingawa, M., Ando, T. and Fry, B.D. (eds) Stable Isotopes in the Biosphere. Kyoto University Press, Kyoto, 78-91.

Ogrinc, N., Košir, I.J., Kocjančić, M. and Kidrič, J. (2001) Determination of authenticity, regional origin, and vintage of Slovenian wines using a combination of IRMS and SNIF-NMR analyses. Journal of Agricultural and Food Chemistry 49, 1432-1440.

O'Leary, M.H. (1993) Biochemical basis of carbon isotope fractionation. In: Ehleringer, J.R., Hall, A.E. and Farqnhar, G.D. (eds) Stable isotopes and plant carbon-water relations. Academic, San Diego, 19-28.

Oliveri, P. and Downey, G. (2012) Multivariate class modeling for the verification of food-authenticity claims. Trends in Analytical Chemistry 35, 74-86.

Organisation Internationale de la Vigne et du Vin (2001) Resolution Oeno 17-2001. Détermination du rapport isotopique du l'éthanol (Office Internationale de la Vigne et du Vin: Paris, France).

Organisation Internationale de la Vigne et du Vin (2009) Resolution Oeno 353-2009. Méthode de détermination du rapport isotopique 180/160 de l'eau dans le vin et le moût (Office Internationale de la Vigne et du Vin: Paris, France).

Organisation Internationale de la Vigne et du Vin (2011) Resolution Oeno 426-2011. Détermination de la repartition du deuterium dans l'éthanol issu de moûts de raisins fermentés, moûts de raisins concentres, moûts de raisins concentres rectifies etde vins par resonance magnétique nucléaire (SNIFNMR/RMNFINS) (Office Internationale de la Vigne et du Vin: Paris, France).

Papotti, G., Bertelli, D., Graziosi,R., Silvestri,M., Bertacchini, L., Durante, C. and Plessi, M. (2013) Application of One- and Two-Dimensional NMR Spectroscopy for the Characterization of Protected Designation of Origin Lambrusco Wines of Modena, Journal of Agricultural and Food Chemistry 61, 1741−1746.

Pereira, L., Guedes-Pinto, H. and Martins-Lopes, P. (2011) An Enhanced Method for *Vitis vinifera* L. DNA Extraction from Wines. American Journal of Enology and Viticulture 62:4, 547-552.

Peres, B., Barlet, N., Loiseau, G. and Montet, D. (2007) Review of the current methods of analytical traceability allowing determination of the origin of foodstuffs. Food Control 18, 228–235.

Petrini, R., Sansone, L., Slejko, F.F., Buccianti, A., Marcuzzo, P. And Tomasi, D. (2015) The 87Sr/86Sr strontium isotopic systematics applied to Glera vineyards: A tracer for the geographical origin of the Prosecco. Food Chemistry 170, 138–144.

Rapeanu, G., Vicol, C. and Bichescu, C. (2009) Possibilities to assess the wines authenticity. Innovative Romanian Food Biotechnology 5, 1- 9.

Romano, A., Fischer, L., Herbig, J., Campbell-Sills, H., Coulon, J., Lucas, P., Cappellin, L. and Biasioli, F. (2014) Wine analysis by FastGC proton-transfer reaction-time-of-flight-mass spectrometry. International Journal of Mass Spectrometry 369, 81–86.

Roßmann, A., Reniero, F., Moussa, I., Schmidt, H.L., Versini, G. and Merle M.H. (1999) Stable oxygen isotope content of water of EU data-bank wines from Italy, France and Germany. Z Lebensm Unters Forsch A 208:400–407.

Roßmann, A., Schmidt, H.L., Reniero, F., Versini, G., Moussa, I., Merle, M.H. (1996) Stable carbon isotope content in ethanol of EC data bank wines from Italy, France and Germany. Z Lebensm Unters Forsch 203, 293-301.

Roßmann, A., Haberhauer, G., Hölzl, S., Horn, P., Pichlmayer, F. and Voerkelius, S. (2000) The potential of multielement stable isotope analysis for regional origin assignment of butter. European Food Research *and* Technology 211, 32–40.

Rozanski, K., Araguás-Araguás L. and Gonfiantini,R. (1993) Isotopic patterns in modern global precipitation. Climate Change in Continental Isotopic Records, Geophys. Monogr. Ser., 78P. R. Swart, et al., 1–36, AGU, Washington D.C., 1993.

Rummel, S., Hoelzl, S., Horn, P., Rossmann, A. and Schlicht, C. (2010) The combination of stable isotope abundance ratios of H, C, N and S with 87Sr/86Sr for geographical origin assignment of orange juices. Food Chemistry 118, 890–900.

Saccenti, E., Hoefsloot, H.C.J, Smilde, A.K., Westerhuis, J.A. and Hendriks, M.W.B. (2014) Reflections on univariate and multivariate analysis of metabolomics data. Metabolomics 10, 361–374.

Santato, A., Bertoldi, D., Perini, M., Camin, F. and Larcher, R. (2012) Using elemental profiles and stable isotopes to trace the origin of green coffee beans on the global market. Journal of Mass Spectrometry 47 (9), 1132-1140.

Scartazza, A., Mata, C., Matteucci, G., Yakir, D., Moscatello, S. and Brugnoli, E. (2004) Comparisons of $\delta^{13}C$ of photosynthetic products and ecosystem respiratory CO_2 and their responses to seasonal climate variability. Oecologia 140, 340-351.

Scheidegger, Y., Saurer, M., Bahn, M. and Siegwolf, R. (2000) Linking stable oxygen and carbon isotopes with stomatal conductance and photosynthetic capacity: a conceptual model. Oecologia 125, 350–357.

Schellenberg, A., Chmielus, S., Schlicht, C., Camin, F., Perini, M., Bontempo, L., Heinrich, K., Kelly, S.D., Roßmann, A., Thomas, F., Jamin, E. and Horacek, M. (2010) Multielement stable isotope ratios (H, C, N, S) of honey from different European regions. Food Chemistry 121, 770–777.

Swoboda, S., Brunner, M., Boulyga, S.F., Galler, P., Horacek, M. and Prohaska, T. (2008) Identification of Marchfeld asparagus using Sr isotope ratio measurements by MC-ICP-MS. Analytical and Bioanalytical Chemistry 390, 487-494.

Taylor, V.F., Longerich, H.P. and Greenough, J.D. (2003) Multielement Analysis of Canadian Wines by Inductively Coupled Plasma Mass Spectrometry (ICP-MS) and Multivariate Statistics. Journal of Agricultural and Food Chemistry 51, 856–860.

This, P., Jung, A., Boccacci, P., Borrego, J., Botta, R., Costantini, L., Crespan, M., Dangl, G.S., Eisenheld, C., Ferreira-Monteiro, F., Grando, M.S., Ibáñez, J., Lacombe, T., Laucou, V., Magalhães, R., Meredith, C.P., Milani, N., Peterlunger, E., Regner, F., Zulini, L. and Maul, E. (2004) Development of a standard set of microsatellite reference alleles for identification of grape cultivars. Theoretical and Applied Genetics 109: 1448-1458.

TRACE TOLL (2009) web application available at http://update.maritech.is/tracetool/publish.htm.

Trincherini, P.R., Baffi, C., Barbero, P., Pizzoglio, E. and Spalla, S. (2014) Precise determination of strontium isotope ratios by TIMS to authenticate tomato geographical origin. Food Chemistry 145, 349–355.

Vasconcelos, A.M.P. and Chaves das Neves, H.J. (1989) Characterization of elementary wines of Vitis vinifera varieties by pattern recognition of free amino acid profiles. Journal of Agricultural and Food Chemistry 37 (4), 931-937.

Versari, A., Laurie, V.F., Ricci, A., Laghi, L. and Parpinello, G.P. (2014) Progress in authentication, typification and traceability of grapes and wines by chemometric approaches. Food Research International 60, 2–18.

Voerkelius, S., Lorenz, G.D., Rummel, S., Quétel, C.R., Heiss, G., Baxter, M., Brach-Papa, C., Deters-Itzelsberger, P., Hoelzl, S., Hoogewerff, J., Ponzevera, E., Van Bocxstaele, M. and Ueckermann, H. (2010) Strontium isotopic signatures of natural mineral waters, the reference to a simple geological map and its potential for authentication of food. Food Chemistry 118, 933–940.

Wachter, H., Christoph, N. and Seifert, S. (2009) Verifying authenticity of wine by Mahalanobis distance and hypothesis testing of stable isotope pattern – a case study using the EU Wine Databank. Mitteilungen Klosterneuburg 59, 237–249.

West, J.B., Ehleringer, J.R. and Cerling, T.E. (2007) Geography and vintage predicted by a novel GIS model of wine $\delta^{18}O$. Journal of Agricultural and Food Chemistry 55, 7075-7083.

Woodcock, T., Downey, G., Kelly J.D. and O'Donnell, C.P. (2007) Geographical Classification of Honey Samples by Near- Infrared Spectroscopy: A Feasibility Study. Journal of Agricultural and Food Chemistry 55, 9128–9134.

Woodcock, T., Downey, G., O'Donnell, C.P. (2009) Near Infrared Spectral Fingerprinting for Confirmation of Claimed PDO Provenance of Honey. Food Chemistry 114, 742-746.

Chapter 2

2 Statistical methods for improving verification of claims of origin for Italian wines based on stable isotope ratios

N. Dordevic, R. Wehrens, G.J. Postma, L. M. C. Buydens, F. Camin
Analytica Chimica Acta (2012), 757, pp 19-25.

Abstract

Wine derives its economic value to a large extent from geographical origin, which has a significant impact on the quality of the wine. According to the food legislation, wines can be without geographical origin (table wine) and wines with origin. Wines with origin must have characteristics which are essentially due to its region of production and must be produced, processed and prepared exclusively within that region.

The development of fast and reliable analytical methods for the assessment of claims of origin is very important. The current official method is based on the measurement of stable isotope ratios of water and alcohol in wine, which are influenced by climatic factors. The results in this paper are based on 5220 Italian wine samples collected in the period 2000-2010. We evaluate the univariate approach underlying the official method to assess claims of origin and propose several new methods to get better geographical discrimination between samples. It is shown that multivariate methods are superior to univariate approaches in that they show increased sensitivity and specificity. In cases where data are non-normally distributed, an approach based on mixture modelling provides additional improvements.

Keywords

H, C and O isotopes; EU Wine Databank; regional origin; univariate; multivariate; mixture modelling

2.1 Introduction

Wine derives its economic value partly from its geographical origin and falsely declaring a certain origin of wine is a relatively common fraud [1] and therefore is important to be able to assess whether claims of origin are valid [2]. According to food legislation wines can be without origin (table wines) or with origin. Wines with origin must be produced, processed, and prepared in a specific region using traditional methods. DOC (Controlled designation of origin) and DOCG (Controlled designation of origin guaranteed) are geographical indications to protect names of Italian wines [3, 4]. IGT (Indicazione geografica tipica) is the Italian national classification scheme created to recognize high quality wines that do not meet requirements for the DOC and DOCG schemes. Currently, Italy has 329 DOC, 70 DOCG and 118 IGT wines. [5]

The official method to assess claims of origin is based on the measurement of the ratio between stable isotopes (that do not decay over the time) of the biologically important elements (H, C, and O) of wine components. The isotopic ratios of these elements (D/H, $^{13}C/^{12}C$ and $^{18}O/^{16}O$) show variations related to climatic factors and to the photosynthetic cycle of the plant [6]. This therefore gives information about the origin of a wine, and is also usable to detect adulteration such as addition of cane and beet sugar, and water. In particular, referring only to the geographical variability of wine grape, D/H and $^{18}O/^{16}O$ in the plant are influenced by the isotopic content of the local groundwater [7, 8] and by the climatic conditions of the pre-harvest period influencing the evapotranspiration processes occurring in plants [9]. The ratio $^{13}C/^{12}C$ in plant material is influenced by several environmental and physiological factors that influence the stomatal conductance and the intercellular and ambient CO_2 concentration, such as relative humidity, temperature, amount of precipitation and water stress [10-12].

A system of official databanks for all wine-producing countries within European Union territory has been set up since 1991 [13, 14]. According to the current EU Regulation [15], every year a number of samples representative of the wine production of each member state are officially collected. The yearly contribution of Italy is 400 samples, originating from different wine-growing regions – in each case, care has been taken to take samples that are representative for the whole region. For each sample about 10 kilos of fresh grapes are harvested, vinified under controlled conditions and finally the resulting wines are analysed. The resulting isotopic ratios are stored in the databank. The current method to assure claims of origin is to apply univariate t tests for all variables in the databank.

The application of multivariate statistics can lead to big improvements, compared to univariate methods, and there is an increase in the number of applications of multivariate statistics in food authentication and traceability using stable-isotope ratios [e.g., 16-20]. For wines, mostly in combination with other analytical approaches, some publications assess geographical origin [21-23]; fraud detection, based on a small subset of the EU databank, has also been described [24].

In this work, we investigate the univariate approach underlying the official method to assess claims of origin, and propose several extensions, based on the 5220 Italian wine samples collected in the period 2000-2010. As a multivariate extension to the t test, the Hotelling T2 test is used, which is based on the Mahalanobis distance. In addition, mixture modelling can be used to improve the description of the data, leading to narrower confidence intervals and hence better classification properties. This method can be applied both in the univariate (t test) and multivariate (Mahalanobis distance) cases. The results show that significant improvements can be obtained.

2.2 Methods

2.2.1 Samples

The samples from the Italian Wine Databank from 2000 to 2010 are considered, originating from 17 Italian regions. The number of samples per year varies from 449 in 2010 to 534 in 2000. For each sample, five variables based on isotopic ratios are available. The ratio between deuterium and hydrogen of the methyl group in alcohol is indicated with $(D/H)_1$, and the corresponding ratio in the alcohol methylene group with $(D/H)_2$. These parameters are determined with nuclear magnetic resonance (NMR) [25]. The R variable represents the ratio of these two variables ($R= 2(D/H)_2 / (D/H)_1$). The other two isotopic ratios in the databank, $^{13}C/^{12}C$ and $^{18}O/^{16}O$, are measured by isotope-ratio mass spectrometry (IRMS). They are presented in delta notation, in which the stable isotope abundance is expressed relative to a standard, V-PDB (Vienna - Pee Dee Belemnite) for $\delta^{13}C$ and V-SMOW (Vienna – Standard Mean Ocean Water) for $\delta^{18}O$. All data have been obtained according to the official methods [26-28]. Yearly variation in weather conditions affects isotopic variables and samples from the same region cannot be compared across different years. Thus, every combination of region and year forms a

separate “class”. Figure 1 shows a box plot indicating the number of samples per class over the years. Although there is some variation, we see that the classes generally contain reasonable numbers of samples. For every class, we can define “typical” ranges for the isotopic variables, which can be used to decide whether new samples can originate from a particular class or not.

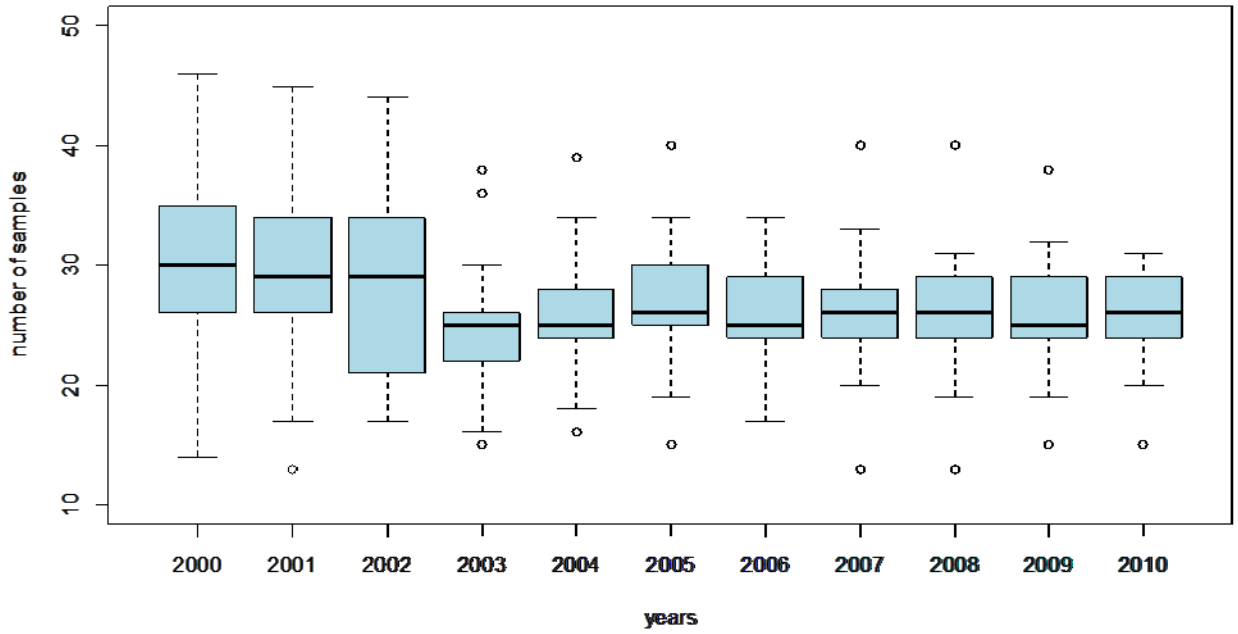

Figure 1. Boxplots of class sizes over all years.

2.2.2 Univariate method

The samples from the Italian Wine Databank can be used to calculate confidence intervals (CI) for the five variables $(D/H)_1$, $(D/H)_2$, R, $\delta^{18}O$, $\delta^{13}C$, assuming normal distributions:

$$CI = \overline{x} \pm t * sd \qquad (1)$$

$\overline{x}$ *- mean value of the examined isotope parameter for the particular class,*

t – t value for the appropriate confidence level and number of degrees of freedom,

sd – standard deviation.

An overall assessment of a claim of origin consists of five t tests, one of every variable. If any of the tests leads to a rejection of the null hypothesis, then the sample is considered to be non-authentic and to originate from another class. In the application of the official method, other considerations are taken into account apart from statistical considerations –

note that five univariate tests at 95% confidence lead to an overall confidence level of 77%, which is very low. For comparison purposes, we focus on the statistical part of the official method and use an overall confidence level of 95%, corresponding to a 99% level of confidence for the individual tests ($0.99^5 = 0.95$). Note that, as in the official method for checking claims of origin, we are treating the five variables as independent, even though one of them (R) is a ratio of two of the others.

2.2.3 The Mahalanobis distance

The multivariate extension of the t test, used in this study, is the Mahalanobis distance (MD) [29]. Rather than doing five separate tests, as in the univariate case, the (squared) Mahalanobis distance measures the distance of a sample to the center of the distribution (given by the values for the wines in a particular class), taking into account the correlation in the data:

$$MD^2 = (X - \overline{X})^T C^{-1} (X - \overline{X}) \qquad (2)$$

X - vector of the isotope data to be examined,

$\overline{X}$ *- mean vector of the isotope data of the particular class,*

C - covariance matrix of the samples from particular class.

The squared Mahalanobis distance approximately follows a chi-squared distribution with m degrees of freedom (m is the number of variables) [30]. As a threshold value, again the 95% quantile of the chi-square distribution is used. Samples which do not exceed the threshold value can be said to belong to the particular class of interest, and they are considered to be authentic. Samples with larger Mahalanobis distances are assessed to not belong to that class, and are considered as non-authentic.

2.2.4 Mixture modelling

Both t tests and Mahalanobis distance assume normally distributed data. In practice, however, it often happens that data are distributed differently: in the Wine Databank, for instance, there are several examples of regions that seem to have multimodal distributions. This can be explained by differences in climate and environmental factors within one region. As a result, confidence intervals are much wider than necessary and statistical tests lose power. Mixture modelling [31] is an approach to describe multimodal

distributions with several gaussians. The optimal number of gaussians is usually determined by criteria like the Bayesian Information Criterion (BIC):

$$BIC = 2\,lnL - n_p \ln n \qquad (3)$$

lnL- log-likelihood

n- number of objects

n_p- total number of parameters to be estimated

The BIC strikes a balance between an accurate description of the data (a high log-likelihood) and a parsimonious description, using as few parameters as possible (hence the second term, which penalizes solutions with many parameters). Also other approaches, such as crossvalidation, could be used for this purpose. The main advantage of using BIC in this case is that all samples are available simultaneously for the definition of the confidence intervals, whereas in crossvalidation only a subset is used. Even though the wine databank is quite large, for individual classes the number of samples can be limited and using the BIC rather than crossvalidatin allows us to use as many as possible in the definition of the confidence limits.

In order to describe non-normally distributed data from a particular class, mixture modelling uses a weighted sum ($\Phi_{i,tot}$) of the densities of the individual gaussians ($\Phi_{i,c}$). The mixture parameter T_c represents the weight, or proportion, of any gaussian c in the mixture:

$$\Phi_{i,tot} = \Sigma\, T_c \Phi_{i,c} \qquad (4)$$

The values for the mixture proportions and the individual gaussians are typically obtained by application of the expectation maximization (EM) algorithm [32]. Here, the MCLUST software, version 3 [33, 34] for R [35] is used. This software also provides the possibility to impose certain restrictions on the gaussian distributions used in the modelling: one can, e.g., require that they all have the same shape, or orientation, or size, or a combination of these restrictions, leading to different models of different complexity [33]. The advantage of this is that simpler models need fewer parameters to be calculated and are less likely to overfit. The BIC value is then used not only to choose the optimal number of gaussians, but also to identify the optimal model.

Once the optimal parameters have been identified by the EM algorithm, a cutoff value for the density needs to be defined. New samples with a density lower than the cutoff can

then be classified as non-authentic, and samples with a higher density as authentic. Since the maximal number of training instances is around 40, the densities of the original samples are not immediately useful for identifying the threshold. However, the threshold is easily obtained by sampling from the fitted density: generating a large number of random samples from this mixture distribution immediately gives us an accurate estimate of, e.g., the 5% density quantile, corresponding with a univariate 95% confidence interval. A similar approach has previously been used in the context of statistical process control [36].

2.2.4.1 Univariate mixture modelling

Firstly, the mixture modelling approach is applied to every variable separately (univariate approach). The BIC is used to determine the optimal model to describe each variable. Both the number of gaussians, and the type of model are optimized. In this case, only two different models are possible – gaussians forced to have the same variance, or gaussians with different variances. If the optimal number of gaussians for a particular variable equals one, the result for this variable is equal to the t test. If more gaussians are needed, results are different. The threshold for each variable separately is obtained as described above, taking into account the fact that we are again combining the result of five individual tests. Since we aim at an overall confidence level of 95%, the individual tests are again performed at 99%.

2.2.4.2 Multivariate mixture modelling

The multivariate case is handled analogously: multivariate class distributions are fitted with a mixture of gaussians. If the BIC indicates that one gaussian suffices, the results are the same as the results of the Mahalanobis distance; if more gaussians are required, the results are different. Since only one test is done in this case, the multivariate method immediately employs the 95% threshold from the (multivariate) densities. For multivariate distributions, as many as fourteen different mixture models are possible: again, the BIC value is used to choose the optimal one [33].

2.2.5 Evaluation of the statistical models

Results are expressed in the familiar terminology of sensitivity and specificity. Sensitivity, also known as the true positive fraction, is the proportion of samples belonging to the class that are accepted as authentic. Since we have set the confidence level of our statistical

tests to 95%, by definition, the sensitivity in all cases should be 95% or a value close to that. Specificity is defined as the proportion of samples outside of a class which are recognized as non-authentic. This is also known as the true negative fraction, and is related to the power of the test: a high power indicates that non-authentic samples are being recognized as such with high probability.

Note that specificity in our framework is estimated without bias: the models are built using only class samples, whereas specificity is evaluated using the samples that are not in the class. On the other hand, sensitivity in the case of mixture modelling may be slightly lower than expected since the BIC criterion chooses the optimal model and number of gaussians. Therefore, a separate validation step is included to make sure that sensitivity values are at the expected level. We achieve this by performing a crossvalidation, only for the fifteen classes with the largest numbers of samples. The crossvalidation is performed by leaving out a random subset of samples, building the model, and testing the model on the left-out samples. This procedure was repeated 1000 times, and the mean sensitivity values were recorded.

2.3 Results and Discussion

2.3.1 Univariate method

Figure 2 shows sensitivities and specificities for data from all regions in the year 2005, according to the univariate method.

Sensitivity values are 100% for many of the regions (e.g. Sardinia, Lazio, Campania) and at the lowest level is 89.5%, for Calabria. The mean sensitivity over all regions in 2005 is 97.9%. Specificity values, however, vary from 20.2%, 22% and 23% for Marche, Campania and Calabria, to 74.7% and 82.9% for the Trentino Alto Adige (TNAA) and Sardinia regions, with a mean specificity value of 45.8% over all regions. Similar figures are obtained for other years. The sensitivity values therefore are quite good, but it seems that specificity values can be improved.

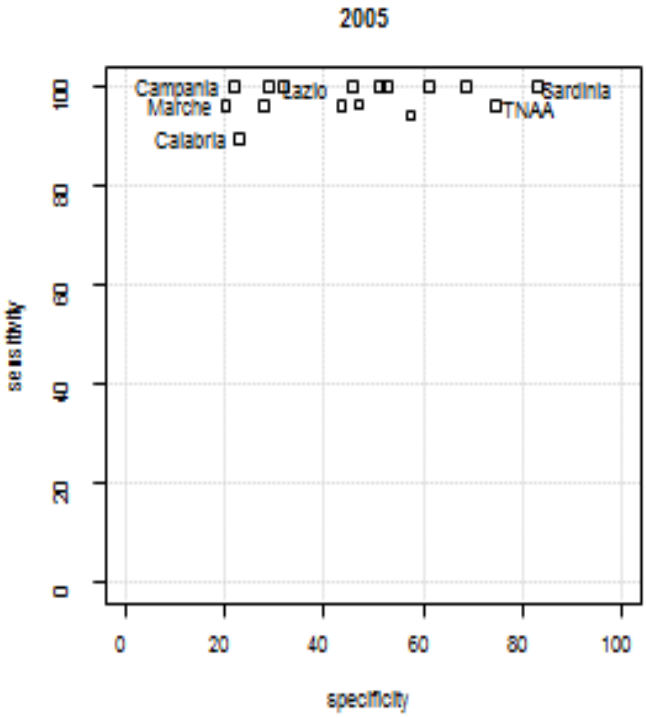

Figure 2. Sensitivity and specificity for all regions from 2005: univariate method. Regions discussed in the text are labeled.

2.3.2 The Mahalanobis distance

Indeed, the other methods proposed in this paper show significant improvements in specificity. Figure 3 presents the example of the Lazio region for all years considered. In all cases, sensitivity values are the same or very similar, so that we focus on specificities only.

For the Mahalanobis distances, the specificity values vary from 41.3% in 2004 to 77.7% in 2008, with a mean value of 55.0%. Since the mean for the univariate method equals 35.3%, using the Mahalanobis distance leads an improvement of almost 20%. Specificity values vary within the mentioned range, but there are big improvements in every single case. A similar pattern can be observed for all other regions. These results indicate that using the Mahalanobis distance, discrimination between samples from different Italian regions is improved dramatically.

As seen in Figure 2, there are large differences between individual regions, depending on their geographic and climatic conditions. The way this affects the Mahalanobis distances can be visualized in Principal Component (PC) space [37]. The positions of samples from Lazio and Trentino Alto Adige are shown in red triangles in Figure 4, whereas data from the other regions are shown in gray dots. Clearly, the samples from Lazio are much more central than the ones from Trentino Alto Adige, and show much more overlap with samples from other regions, the main reason why specificity for Trentino Alto Adige is much higher than for Lazio.

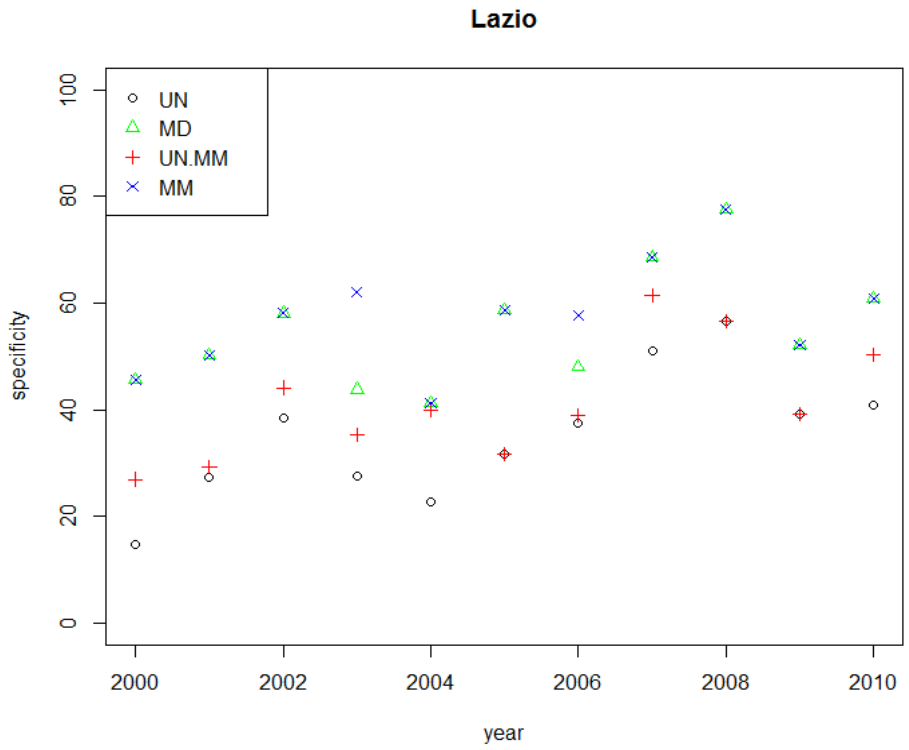

Figure 3. Specificity values of the four different methods for Lazio region from 2000 to 2010. Abbreviations UN, MD, UN.MM and MM correspond to univariate method, the Mahalanobis distance, univariate and multivariate mixture modelling.

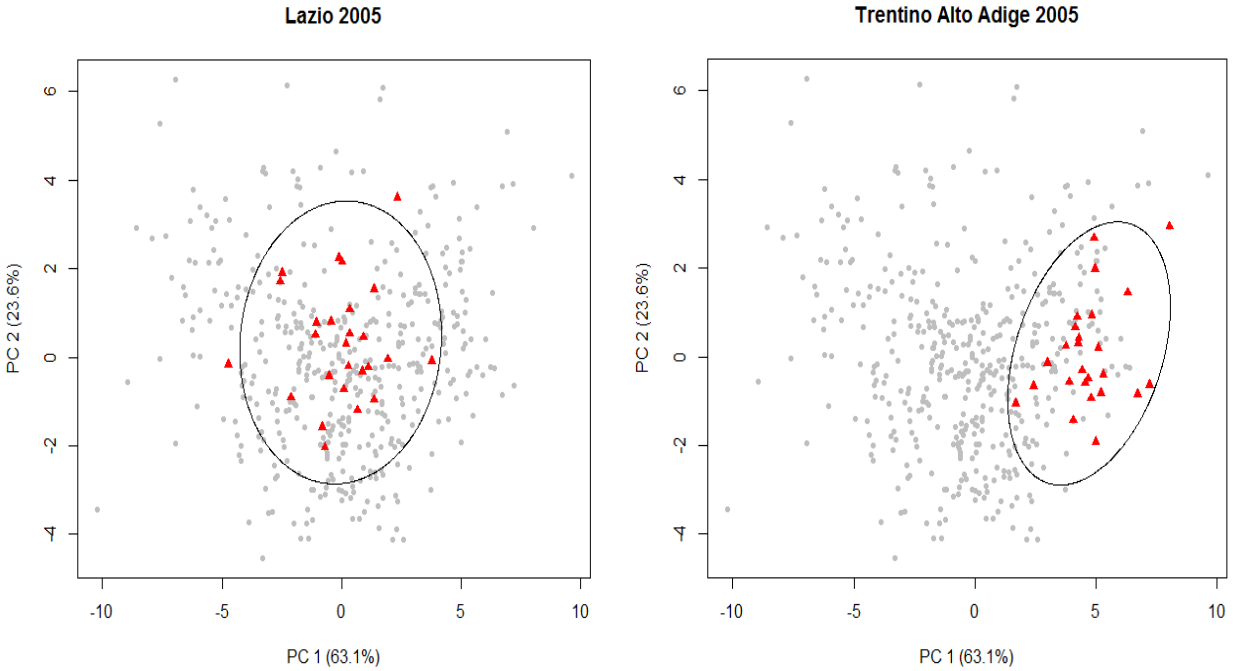

Figure 4. Visualization of Lazio (left) and Trentino Alto Adige (right) reference samples (vintage 2005) using 2 principal components. The ellipses correspond to the value of 95% quantile of the chi-square distribution.

2.3.3 Mixture modelling

2.3.3.1 Mixture modelling validation

To validate the sensitivity results of the BIC-selected models, crossvalidation was performed for the samples from the fifteen biggest classes. Only the biggest classes are considered in order to minimize the bias introduced by the subsampling in the crossvalidation procedures. The results for multivariate mixture modelling are shown in Figure 5, containing the averaged sensitivities. Clearly, sensitivity values drop when using more complex models, i.e. models with more gaussians or with more parameters. The three most complex models, EEV, VEV and VVV lead to quite big decreases in sensitivity, especially for models with more than one gaussian, and these have been excluded from the list. For all other models with one or two gaussians, sensitivity drops are small, and therefore we are considering these seven models with up to two gaussians in the remainder of the paper. For univariate mixture modelling, similar results are obtained and all models with one or two gaussians are allowed. The more complex models excluded in this paper, should only be considered if many more samples would be available.

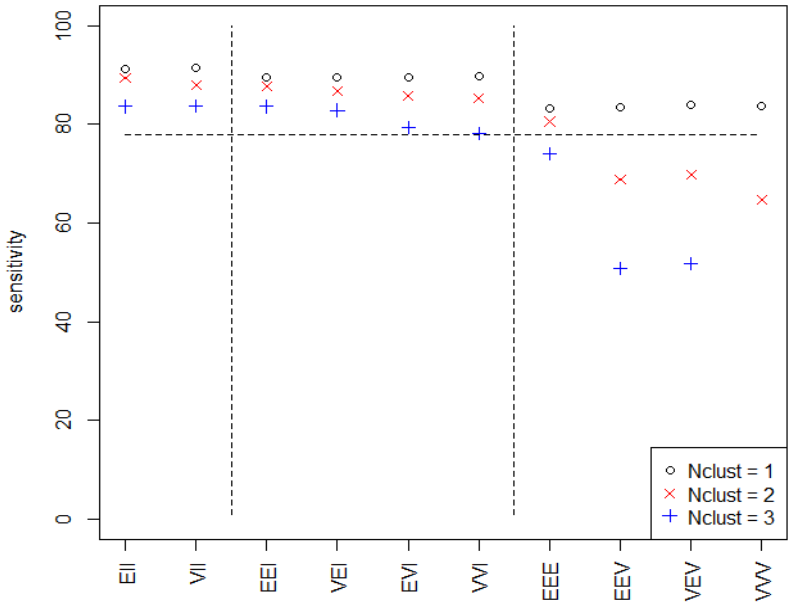

Figure 5. Visualization of average sensitivity values (15 largest classes only) from cross-validation for different multivariate mixture models in terms of shape, volume and orientation [33]. Nclust specifies the numbers of mixture components (clusters).Models are sorted from the simplest model on the left to the most complex model on the right. The horizontal dashed line represents the expected sensitivity from a univariate model testing at $0.95^5 = 0.77$, the basis of the current official method. The vertical dashed lines split models with spherical, diagonal and elliptical distributions [33].

2.3.3.2 Univariate mixture modelling

As for the normal univariate model, the confidence level was set to 99% for each variable, leading to an overall confidence level (and hence sensitivity) of 95%. Considering only models with up to two gaussians, Figure 3 shows that the average specificity for Lazio increases from 35.3% (univariate method) to 41.1%, which is an improvement of almost 6%. In the years 2005, 2008 and 2009, no improvement is obtained: these are the cases where one gaussian is optimal, and then the results are equal by definition. In cases where two gaussians are used, improvements are obvious. For example, the specificity for Lazio samples in 2004 increases from 22.7% to 39.9%; the specificity for 2000 increases from 14.9% to 27.0%.

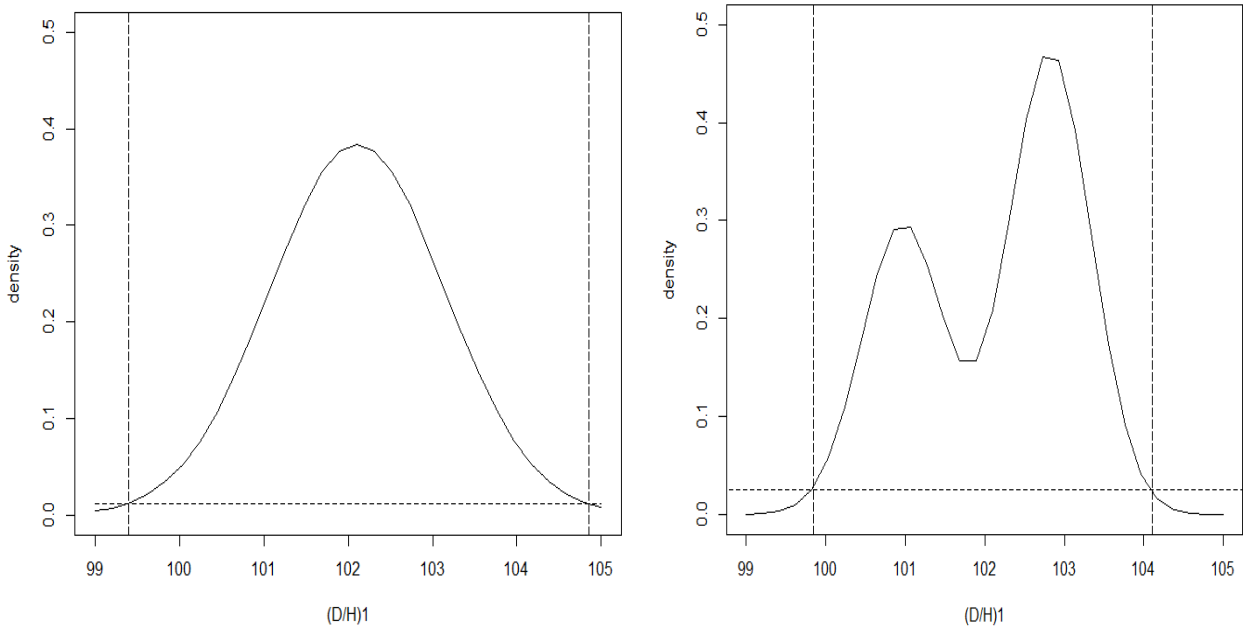

Figure 6. Visualization of distribution of $(D/H)_1$ for Campania 2005 samples defined with one (left) and two (right) Gaussians. The vertical dashed lines correspond to threshold values, and horizontal dashed line to the 99% percentile of density. The confidence interval on the right is narrower than that on the left.

The reason for the improvement is that univariate mixture modelling creates narrower confidence intervals for non-normally distributed parameters, resulting in less overlap with the samples of other regions (Figure 6). Assuming the normal distribution, the 99% confidence level for $(D/H)_1$ in Campania 2005 samples range from 99.4 to 104.9; corresponding interval for the two-gaussians solution is from 99.8 to 104.1.

2.3.3.3 Multivariate mixture modelling

In the multivariate mixture modelling results, only models with up to two gaussians are considered, and the three most complex models are excluded. The most frequently selected model is defined with one gaussian. In multidimensional space these models have the same shape and orientation as model obtained by Mahalanobis distance and the results are the same. This is exactly analogous to the univariate case.

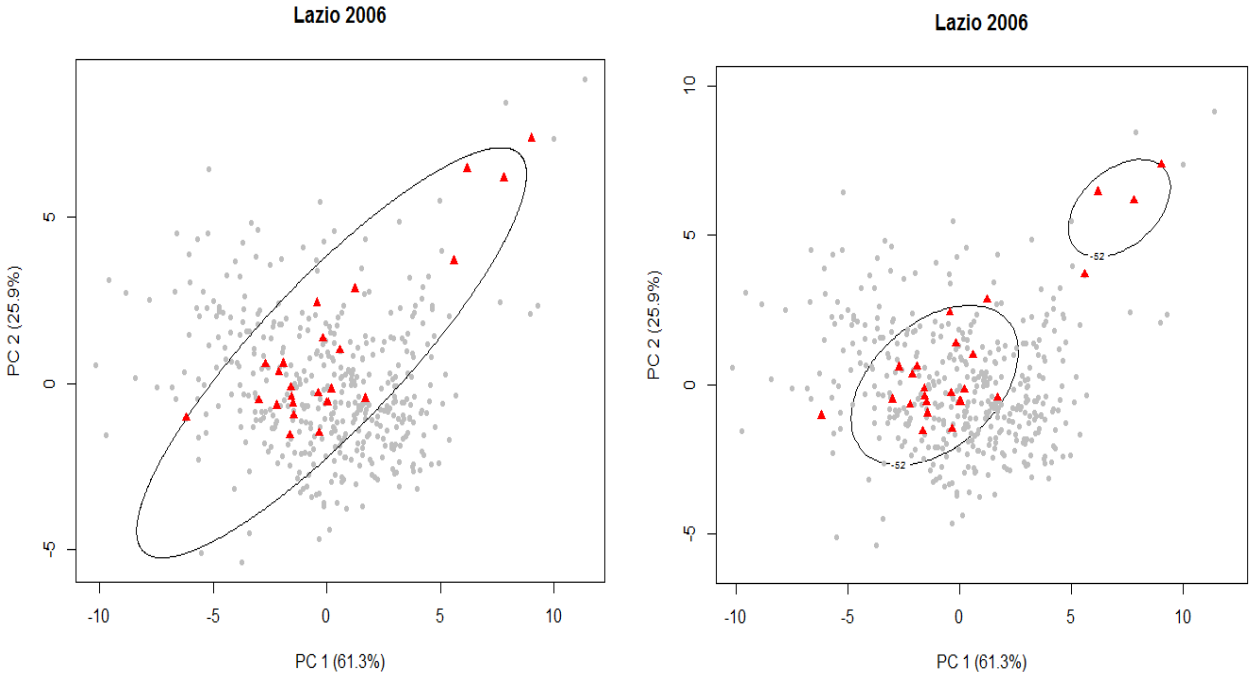

Figure 7. Visualization of Lazio reference samples from 2006 using 2 principal components. The left plot shows the one-gaussian density (corresponding to the Mahalanobis distance criterion), and the right plot shows the two-gaussian mixture modelling approach. The smaller volume covered with the mixture modelling solution leads to an increase of the specificity.

In cases where two gaussians are selected, this results in better discrimination between samples from different classes, i.e., higher specificity values. Figure 3 shows that in two years, a significant improvement is obtained over the Mahalanobis distance by using mixture models with two gaussians. Specificity values for Lazio samples reach 62.0% in 2003 and 57.8% in 2006, while with the Mahalanobis distance those values are 43.8% and 48.1%, respectively. Also in other classes the same behavior can be observed. For 2004 and 2010, five regions are defined with two gaussians; for 2006, four regions; for 2001 and 2002, three regions; in 2007, 2008 and 2009, two regions are defined with two gaussians and in 2000, 2003 and 2005 only one. The most frequent region defined with two gaussians is Umbria, which indicates the existence of heterogeneity within that region.

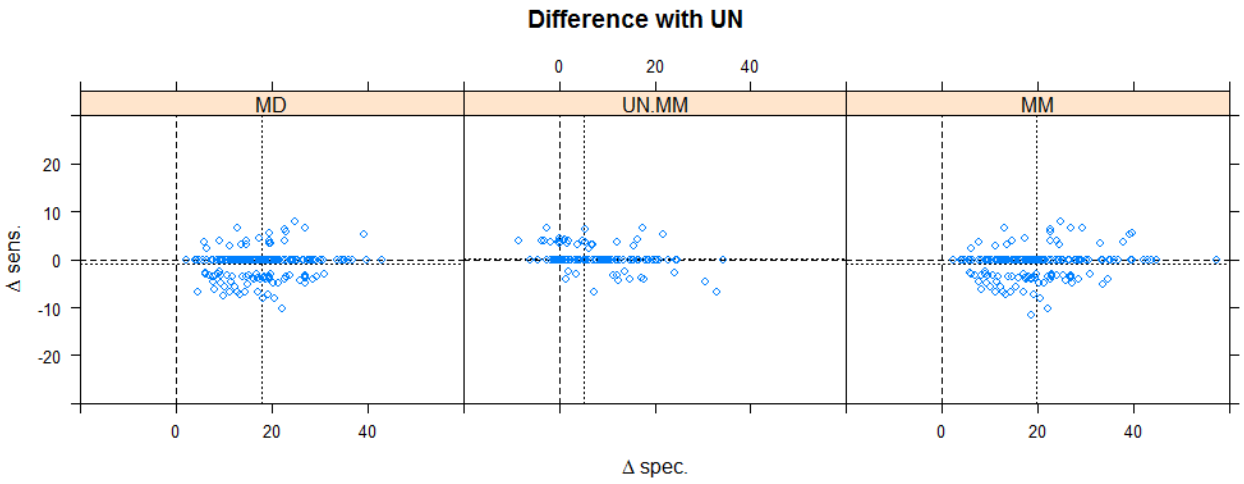

Figure 8. Differences in sensitivities and specificities between different methods for analysed regions from 2000 to 2010, compared to the univariate method. Dotted lines represent average values of sensitivity and specificity differences.

The improvements in specificity upon using more than one gaussian are illustrated in PC space for samples for Lazio, 2006 (Figure 7). The data are not normally distributed and when defined with a mixture of 2 gaussians, occupy less volume in multidimensional or PC space than the corresponding one-gaussian model.

Summarizing, multivariate methods are better than univariate methods, and additional improvements can be obtained when the data are not well described by one normal distribution. This is visualized in Figure 8, showing the differences between the three extensions proposed in this paper and the univariate method. The improvements in specificity are substantial for all three (the blue dots are to the right of the origin) except for few cases for univariate mixture modelling. The average increase in specificity is 5% using the univariate mixture modelling, 18.0% for the Mahalanobis distance, and 19.8% for the multivariate mixture modelling (indicated with the vertical dotted lines in the Figure). On the order hand, the effect on sensitivity is small: the averages are very close to zero, indicating that the methods lose nothing in comparison to the univariate method.

2.4 Conclusion

Claims of authenticity have profound economic consequences, and therefore it is becoming more and more important to be able to evaluate these claims. This is especially true for relatively high-prized commodities like wine. This paper expands on the current practice of using an ensemble of five univariate t tests, providing three more powerful alternatives. In particular, multivariate methods provide big improvements in power, so that false claims can be discovered much more easily. Taking into account heterogeneity in climatic and geographic conditions in wine-growing areas, using mixture modelling approaches, leads to further improvements. For this purpose, the existence of a large databank of representative samples is indispensible. In the future, additional improvements in authenticity evaluation can be expected if the number of samples per region and per year could be improved, or if additional informative variables could be measured.

References

[1] L. Holmberg, Int. J. of Wine Res. 2 (2010) 105–113.

[2] S. Kelly, K. Heaton, J. Hoogewerff, Trends Food Sci. Technol. 16 (2005) 555-567.

[3] European Commission, European Union (EU) Agricultural Product.
Quality Policy, Specific EU Quality Schemes – Guaranteeing Quality (http://ec.europa.eu/agriculture/quality)

[4] E.C. Regulation No. 510/06 on the Protection of Geographical Indications and Designations of Origin for Agricultural Products and Foodstuffs.

[5] Ministero delle politiche agricole alimentari e forestali, Dipartimento delle politiche competitive del mondo rurale e della qualità. Direzione Generale dello sviluppo agroalimentare e della qualità. SAQ IX. NUMERO DEI VINI A D.O.C.G. - D.O.C. - I.G.T. AL 19 OTTOBRE 2011.

[6] N. Christoph, A. Rossmann, C. Schlicht, S. Voerkelius, in S. E. Ebeler, G. R. Takeoka, P. Winterhalter (Eds.), Authentication of Food and Wine, American Chemical Society, Washington, 2006, pp. 166-179.

[7] I. Clark, P. Fritz, Environmental Isotopes in Hydrogeology, CRC Press/Lewis Publishers, Boca Raton, FL, 1997.

[8] G. Van der Veer, S. Voerkelius, G. Lorentz, G. Heiss, J.A. Hoogewerff, J. Geochem. Explor. 101 (2009) 175-184.

[9] A. Rossmann, F. Reniero, I. Moussa, H.-L. Schmidt, G. Versini, M.H. Merle, Z. Lebensm. Unters. Forsch. 208 (1999) 400–407.

[10] M.H. O'Leary, in: E. Wada, T. Yoneyama, M. Mingawa, T. Ando, B.D. Fry (Eds.), Stable Isotopes in the Biosphere, Kyoto University Press, Kyoto, 1995, pp. 78-91.

[11] Y. Scheidegger, M. Saurer, M. Bahn, R. Siegwolf, Oecologia 125 (2000) 350–357.

[12] G.D. Farquhar, M.H. O'Leary, J.A. Berry, Aust. J. Plant Physiol. 9 (1982) 121-37.

[13] E.E.C. Regulation No. 2347/91 on the collection of samples of wine products for the purposes of cooperation between Member States and for analysis by nuclear magnetic resonance, including analysis for the purposes of the Community databank.

[14] E.E.C. Regulation No. 2348/91 establishing a databank for the results of analyses of wine products by nuclear magnetic resonance of deuterium.

[15] E.C. Regulation No. 555/2008 laying down detailed rules for implementing Council Regulation (EC) No 479/2008 on the common organisation of the market in wine as regards support programmes, trade with third countries, production potential and on controls in the wine sector.

[16] F. Angerosa, O. Breas, S. Contento, C. Guillou, F. Reniero, E. Sada, J. Agric. Food Chem. 47 (1999) 1013-1017.

[17] F. Camin, L. Bontempo, K. Heinrich, M. Horacek, S. D. Kelly, C. Schlicht, F. Thomas, F. J. Monahan, J. Hoogewerff, A. Roßmann, Anal. Bioanal. Chem. 389 (2007) 309–320.

[18] A. Schellenberg , S. Chmielus, C. Schlicht, F. Camin, M. Perini, L. Bontempo, K. Heinrich, S. D. Kelly, A. Roßmann, F. Thomas, E. Jamin, M. Horacek, Food Chem. 121 (2010) 770–777.

[19] L. Bontempo, F. Camin, L. Manzocco, G. Nicolini, R. Wehrens, L. Ziller, R. Larcher, Rapid Commun. Mass Spectrom. 25 (2011) 899–909.

[20] F. Camin, R. Wehrens, D. Bertoldi, L. Bontempo, L. Ziller, M. Perini, G. Nicolini, M.Nocetti, R. Larcher, Anal. Chimica Acta 711 (2012) 54– 59.

[21] N. Ogrinc, I. J. Košir, M. Kocjančić, J. Kidrič, J. Agric. Food Chem. 49 (2001) 1432-1440.

[22] G. Gremaud, S. Quaile, U. Piantini, E. Pfammatter, C. Corvi, Eur. Food Res. Technol. 219 (2004) 97–104.

[23] X. Capron, J. Smeyers-Verbeke, D.L. Massart, Food Chem. 101 (2007) 1585–1597.

[24] H. Wachter, N. Christoph, S. Seifert, Mitt. Klosterneubg. 59 (2009) 237-246.

[25] E.E.C. Regulation No. 2676/90 determining Community methods for the analysis of wines.

[26] Method OIV MA-AS2-12. Method for $^{18}O/^{16}O$ isotope ratio determination of water in wines and must.

[27] Method OIV MA-AS312-06. Determination by isotope ratio mass spectometry $^{13}C/^{12}C$ of wine ethanol or that obtained through the fermentation of musts, concentrated musts or grape sugar.

[28] Method OIV-OENO 426-2011. Determination of the deuterium distribution in ethanol derived from fermentation of grape musts, concentrated grape musts, rettified concentrates grape musts and wines by application of nuclear magnetic resonance (SNIFNMR/RMNFINS).

[29] R. De Maesschalck, D. Jouan-Rimbaud, D.L. Massart, Chemometr. Intell. Lab. 50 (2000) 1–18.

[30] K. Varmuza, P. Filzmoser, Introduction to Multivariate Statistical Analysis in Chemometrics, Taylor & Francis - CRC Press, Boca Raton, FL, 2009.

[31] G. McLachlan, D. Peel, Finite Mixture Models, Wiley, New York, 2000.

[32] A. P. Dempster, N. M. Laird, D. B. Rubin, J. R. Stat. Soc. Series B (Methodological), 39 (1) (1977) 1-38.

[33] C. Fraley and A. E. Raftery, J. Classif. 20 (2003) 263-286.

[34] C. Fraley and A. E. Raftery, JASA 97 (2002) 611-631.

[35] R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

[36] U. Thissen, H. Swierenga, A. P. de Weijer, R. Wehrens, W. J. Melssen, L. M. C. Buydens, J. Chemometr. 19 (2005) 23–31.

[37] I.T. Jolliffe, Principal Component Analysis, Second Edition, Springer Series in Statistics, Springer, New York, 2002.

Chapter 3

3 Detecting the addition of sugar and water to wine

N. Dordevic, F. Camin, R. M. Marianella, G.J. Postma, L. M. C. Buydens, R. Wehrens Australian Journal of Grape and Wine Research (2013), 19, pp 324-330.

Abstract

Background and Aims: Wine fraud has been practised since ancient times, continuing up to the present day and taking many forms. In this study, detection of the addition of unpermitted sugar and water to wines was evaluated using a univariate method, officially adopted, and several more effective multivariate methods.

Methods and Results: The results are based on 5220 Italian wine samples belonging to the European Union (EU) Wine Databank collected in the period 2000–2010. This databank contains stable isotope ratios for wine samples from many regions over many years. The addition of variable amounts of sugar and water was simulated according to experimental results, and univariate and multivariate methods were compared. It was shown that this type of fraud can be discovered much more easily by the application of multivariate methods.

Conclusions: It can be difficult to detect the addition of unpermitted sugar and water to wines. With multivariate tests based on the EU Wine Databank, the likelihood of detecting this kind of fraud is significantly increased.

Significance of the Study: This research evaluates and proposes a more efficient method for statistical interpretation of the EU Wine Databank for evaluation of wine authenticity.

Keywords: *adulteration*, *EU Wine Databank*, *stable isotope ratio*, *wine authenticity*

3.1 Introduction

Evaluation of the authenticity of wine involves checking several claims, such as declaration of origin, vintage and production method. In the wine industry, there are many examples of producers using fraud in order to increase profit (Holmberg 2010). Beet sugar, cane sugar or rectified must can be added to grape must or wine before or during fermentation to increase the natural ethanol content and therefore the value of the wine. This procedure is legal only for specific wine-growing regions and vintages that suffer from a lack of sun hours during the grape-ripening period, leading to low grape sugar content. Wine is unstable without a certain content of alcohol, and the addition of sugar before fermentation may be vital for wine stability. For other areas, including the whole of Italy, the addition of sugar is not allowed. Dilution of wine with water is also considered a fraud in the wine sector. According to European Regulations 479/2008 and 1234/2007 (European Commission 2007, 2008a), wine is defined as a product obtained exclusively from the alcoholic fermentation of fresh grapes, whether crushed or not, or grape must. In this paper, the addition of both sugar and water will be described using the term 'adulteration'. In order to prevent such fraud, the development of fast and reliable analytical methods for fraud detection is important.

Measurement of stable isotope ratios of the biologically important elements (H, C and O) for grape products has been adopted as the official method by European Union (EU) regulations and by the Organisation Internationale de la Vigne et du Vin in order to detect the addition of water and sugar to grape products, based on comparison of sample data with that obtained for authentic products with the same origin. This has led to the creation of the European Wine Databank (European Commission 1990, 1991a,b, 1997, 2003, 2008b).

In Italy, at least 400 reference samples are included in this Wine Databank every year (collecting a specific number of samples from each individual region). The five stable isotope ratios $(D/H)_1$, $(D/H)_2$, R, $\delta^{18}O$ and $\delta^{13}C$ are used to calculate confidence limits for authentic samples. Authenticity assessment then consists of five one-sample *t*-tests. If even one variable of a wine sample under investigation falls outside those limits, this sample is considered as nonauthentic. Note that such an approach is independent of the type of adulteration.

In the literature, there are few studies of the application of multivariate statistics in wine authentication using stable isotope ratios (Ogrinc et al. 2001, Gremaud et al. 2004, Capron

et al. 2007, Wachter et al. 2009). Multivariate interpretation of the Italian Wine Databank has already shown that it is possible to improve differentiation between samples originating from different regions as compared with a univariate approach (Dordevic et al. 2012). Here, we focus on detecting the addition of sugar and water to wine, comparing the univariate approach based on five individual *t*-tests with the multivariate extension (the Hotelling T2 test) which uses the Mahalanobis distance. The analysis is based on simulation of various quantities of sugar and water added to wine.

3.2 Materials and methods

3.2.1 Data

In this study, the isotopic values of samples from the 17 largest Italian regions (in terms of number of samples) from the Italian Wine Databank for the period 2000–2010 were considered. These samples represent the full range of diversity within one region, in terms of climate, vineyard location, grape cultivar and date of harvest. The five variables were obtained according to official methods (Organisation Internationale de la Vigne et du Vin 2001, 2009, 2011). $(D/H)_1$ represents the ratio between deuterium and hydrogen of the methyl group in alcohol and $(D/H)_2$ is the corresponding ratio in the alcohol methylene group. The ratio of these two variables $[2(D/H)_2/(D/H)_1]$ is the third variable R. The $^{13}C/^{12}C$ and $^{18}O/^{16}O$ variables are expressed in delta notation, in which stable isotope abundance is expressed relative to a standard, Vienna – PeeDee Belemnite for $\delta^{13}C$ and Vienna – Standard Mean Ocean Water for $\delta^{18}O$.

These isotopic ratios of wine show variations related to climatic factors in the preharvest period, which can change between years. Every combination of region and year is considered independently and forms a separate ‘class’. Samples from one class are used to define ‘typical’ ranges for the isotopic variables, which are used to evaluate the authenticity of wines on the market with the corresponding origin and vintage stated on the label.

3.2.2 Fraud sample simulation

Natural ratios of $^{13}C/^{12}C$ and $^{2}H/^{1}H$ isotopes in plant organic materials differ because of isotopic fractionation occurring during photosynthesis. For instance, carbon isotopic ratios

are influenced by the specific pathways active during carboxylation: C_4 plants, such as maize and sugar cane, use the enzyme phosphoenol pyruvate-carboxylase (Hatch-Slack pathway) while C_3 plants (wheat, grapes and sugar beet) fix CO_2 directly through the enzyme ribulose-biphosphate-carboxylase (Calvin pathway). This means that grape and beet sugar can be distinguished from cane sugar on the basis of this isotopic ratio.

The $(D/H)_1$ ratio of ethanol from beet sugar is around 92.5 ppm, which is significantly lower than that of ethanol from wines or fruits (Martin et al. 1988). Addition of beet sugar to grapes therefore significantly decreases $(D/H)_1$ in proportion to the amount of added beet sugar. The $(D/H)_2$ value should not change significantly because it derives mainly from fermentation water. Furthermore, the $\delta^{13}C$ value is not affected significantly because sugar beet belongs to the group of C_3 plants, and the $\delta^{13}C$ value of ethanol from C_3 plants is in the same range. The $(D/H)_1$ ratio of ethanol from cane sugar is around 110 ppm, which is significantly higher than that from wine ethanol. Furthermore, the $\delta^{13}C$ value of cane ethanol is significantly higher than that in wine ethanol (Roßmann et al. 1996, Christoph et al. 2003). This is confirmed by van Jaarsveld (2008), who showed that the addition of cane sugar to grape must leads to a significant linear change in the $(D/H)_1$ and $\delta^{13}C$ values of wine ethanol in relation to the amount of added sugar, whereas it does not lead to any significant change in $(D/H)_2$ and $\delta^{18}O$ values. Dilution with water shows little or no effect on the $(D/H)_1$, $(D/H)_2$ and $\delta^{13}C$ values of wine ethanol, but causes a significant linear change in the $\delta^{18}O$ value of wine (van Jaarsveld 2008).

To assess the potential of different fraud detection methods, we used these linear relationships to create simulated fraud samples. In order to obtain a large number of samples for accurate evaluation of the results, 1000 new samples were generated by drawing from multivariate normal distributions with the means and covariances given by the experimental data for the individual classes. These simulated, unadulterated samples were then used as a base to simulate the addition of different proportions of sugar (both cane and beet) and dilution with water. For instance, the proportion of added sugar is given by the following equation where $m_{added\ sugar}$ and $m_{grape\ sugar}$ represent the amount of added and original grape sugar:

$$\text{e.g.} \quad (m_{added\ sugar} / (m_{added\ sugar} + m_{grape\ sugar})) * 100 \quad (\%) \qquad (1)$$

The mean isotopic values used for cane and beet sugar are from Christoph et al. (2003) and are summarised in Table 1. In comparison with the isotopic values for sugar, the

isotopic value of water is less constant and varies in relation to the geographic and climatic conditions of origin (van der Veer et al. 2009). According to Longinelli and Selmo (2003), mean oxygen isotopic precipitation values in Italy, which are analogous to $\delta^{18}O$ values for ground and tap water, vary from approximately −9‰ in the north and central inland areas to −5‰ in the south (excluding locations higher than 1500 m above sea level. Because of the varying $\delta^{18}O$ values of ground water, these extremes were the two values used in the simulations (Table 1). In this study, increasing fractions of added water, cane and beet sugar (from 1 to 100%) were simulated. It was assumed that 100% of sugar is fermented during the winemaking process. Thus, a series of adulterated samples was obtained for each of the 1000 simulated unadulterated samples (from 1 to 100% adulteration). It should be noted that even 100% adulteration, meaning that no grapes at all were used to produce the wine, can be encountered in practice.

Table 1. Typical values for the isotopic ratios of wines (95% confidence interval) and the values for tap water, beet and cane sugar used for fraud simulation.

	$(D/H)_1$	$(D/H)_2$	$\delta^{13}C$	$\delta^{18}O$
Wines, 95% CI	98.8 to 106	124.5 to 135.5	-29.3 to -24.3	-1.3 to 8.9
Beet sugar	92.5		-27.5	
Cane sugar	109.5		-12	
Water (North Italy)				-9
Water (South Italy)				-5

Empty cells in the table indicate that this particular adulteration did not significantly affect the corresponding isotopic ratio and was not modelled in our simulations.CI, confidence interval.

To validate the results from fraud sample simulations, adulteration of wine samples was performed in our laboratory by adding different amounts of beet and cane sugar to the must before fermentation and by dilution with water. Then, isotopic parameters were measured. In particular, an increasing amount of crystalline beet and cane sugar (seven and six samples, respectively, with an increasing concentration of sugar, from 0 to 180 g/L was added to a fresh must (15°Brix) and subjected to complete fermentation with *Saccharomyces cerevisiae* yeast (Lallemand Inc., Montréal, QC, Canada), following the official method for must (Organisation Internationale de la Vigne et du Vin 2011). An

increasing amount of tap water, from 20 to 40%, was added to one sample of wine. The ethanol distillation procedure and the analysis of the isotope ratios of ethanol and wine water followed the official methods (Organisation Internationale de la Vigne et du Vin 2001, 2009, 2011).

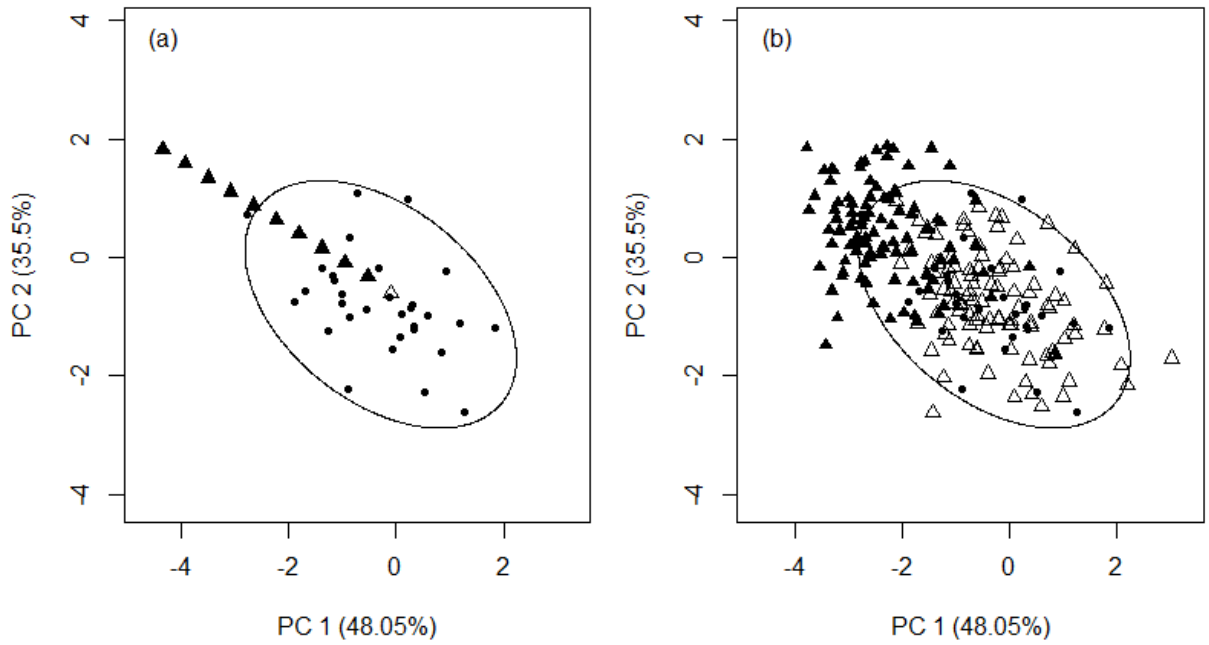

Figure 1. Position of simulated samples for Tuscany in the space of the first two principal components, based on all experimental data, 2007 vintage. (●) Tuscan reference samples; the contour line corresponds to the 95% confidence level of the chi-squared distribution for these samples. (a) One simulated sample (△) from this distribution has been used to simulate the addition of an increasing amount of (▲) cane sugar (from 5 to 50% in 5% intervals). (b) A large number of samples has been simulated according to (△) the same distribution, (▲) along with the corresponding fraud samples with addition of 25% cane sugar. PC, principal component.

3.2.3 Method evaluation

The results are expressed in terms of two parameters: detection level and detection rate. The detection level represents addition of the lowest proportion of sugar (cane and beet) or water, which can be detected by a particular method. Put differently, it is the adulteration level at which a simulated fraud sample exceeds the confidence interval (CI) for that particular class. Mean detection levels can be calculated for both univariate and multivariate methods: the lower these values, the better.

The detection rate is the proportion of fraud samples detected by a particular method at a certain adulteration level. If the detection rate for a particular adulteration level is 100%, this means that all such adulterated samples will be detected. Again, the 1000 simulated

unadulterated samples and their adulterated counterparts can be used to evaluate detection rates for the univariate and multivariate methods.

In Figure 1, an example is given for simulations for Tuscany, 2007, where Figure 1a shows the reference Tuscan samples as filled black circles. The contour line in both plots visualises the region where one would normally expect to find samples from Tuscany, 2007, based on the distribution of the reference Tuscan samples. Figure 1a shows the effect of increasing adulteration of one (simulated) sample with cane sugar. It moves away from the original position in a straight line. At some point, it will be outside the ellipse, at which point it will be detected as a fraud sample. This corresponds to the detection level for this particular sample. Similarly, Figure 1b shows many simulated samples with the same adulteration level (black filled triangles), originating from unadulterated simulated samples (open triangles). The proportion of filled triangles outside the ellipse is an indication of the detection rate. Samples are visualized in the principal component (PC) space (Jolliffe 2002).

3.2.4 Univariate method

The univariate method for detecting adulteration in wine samples consists of the repeated application of one-sample *t*-tests, comparing commercial samples with the databank values for the reference samples of a particular class. Normal distribution is assumed to define CIs for all parameters:

$$CI = \overline{x} \pm t * sd \qquad (2)$$

$\overline{x}$ *- mean value of the isotope parameter examined for the particular class,*

t – t value for the appropriate confidence level and number of degrees of freedom,

sd – standard deviation.

Overall evaluation of authenticity is based on five *t*-tests, for all five isotope ratios. The sample is considered as nonauthentic if even one of the *t*-tests rejects the null hypothesis, which means that at least one of the variables falls outside the corresponding CI. To allow for direct comparison between univariate and multivariate methods, the overall confidence level was adjusted to 95%, setting each individual *t*-test to 99% ($0.99^5 = 0.95$) (Dordevic et

al. 2012). The results of the univariate method are indicated with the label ‘UN’ in the text and the figures.

3.2.5 Multivariate method

The Mahalanobis distance (De Maesschalck et al. 2000) is the multivariate extension of the *t*-tests. Instead of doing separate *t*-tests, the squared Mahalanobis distance measures the distance from the centre of distribution of samples from the corresponding class, taking into account the correlation between variables:

$$MD^2 = (X - \overline{X})^T C^{-1} (X - \overline{X}) \qquad (3)$$

X - vector of the isotope data to be examined,
$\overline{X}$ *- mean vector of the isotope data of a particular class,*
C - covariance matrix of samples from a particular class.

In order to obtain 95% confidence as a threshold value, 95% of the chi-squared distribution is used with m degrees of freedom, where m is the number of variables (Varmuza and Filzmoser 2009). Nonauthentic samples are samples in which the Mahalanobis distance from the centre exceeds the threshold. For the results of the multivariate method, the label ‘MD’ is used.

3.3 Results and discussion

Isotopic ratios measured in samples with the addition of different proportions of beet and cane sugar to the must before fermentation and dilution with water after fermentation were used to confirm the validity of our simulations. These were based on literature values, and in Figure 2 one can see that there is excellent agreement between the experimental results (open blue circles) and the parameters used for the simulation (solid red lines). For beet sugar addition, a limited effect was seen in the experimental values for $\delta^{18}O$, but this was attributed to water present in the crystalline sugar (added before fermentation) and was not taken into account in the simulations. Furthermore, there is a slightly different slope for $\delta^{13}C$, depending on the amount of cane sugar added. This of course depends on

the exact values for the sugar used, but clearly the literature values adopted here lead to realistic values.

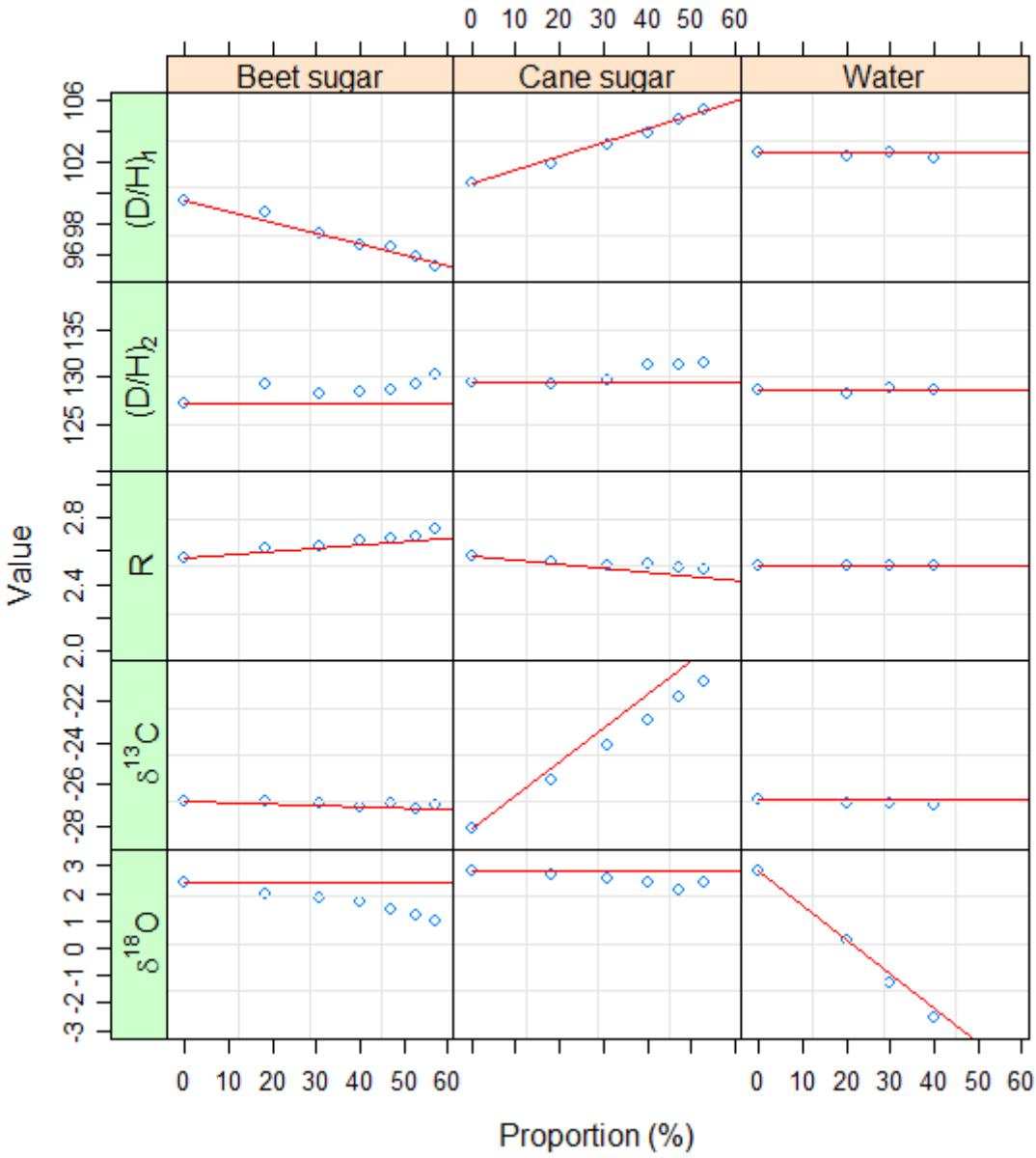

Figure 2. Visualisation of (o) experimental results and (--) simulated isotopic variables due to the increase in the proportion of beet and cane sugar added to grape must before fermentation and tap water added to the wine.

Figure 3 shows the respective results for the 17 Italian regions from 2000 to 2010 in terms of the mean detection level for adulteration with beet and cane sugar, along with two values for water. It can be seen that the multivariate approach is more sensitive. The overall mean detection level for beet sugar was 33.2 and 27.2%, respectively, for the univariate and multivariate methods. For cane sugar, these values were 19.1 and 16.6%,

and in the two simulations of dilution with water, we can also see that the multivariate method obtained a lower mean detection level.

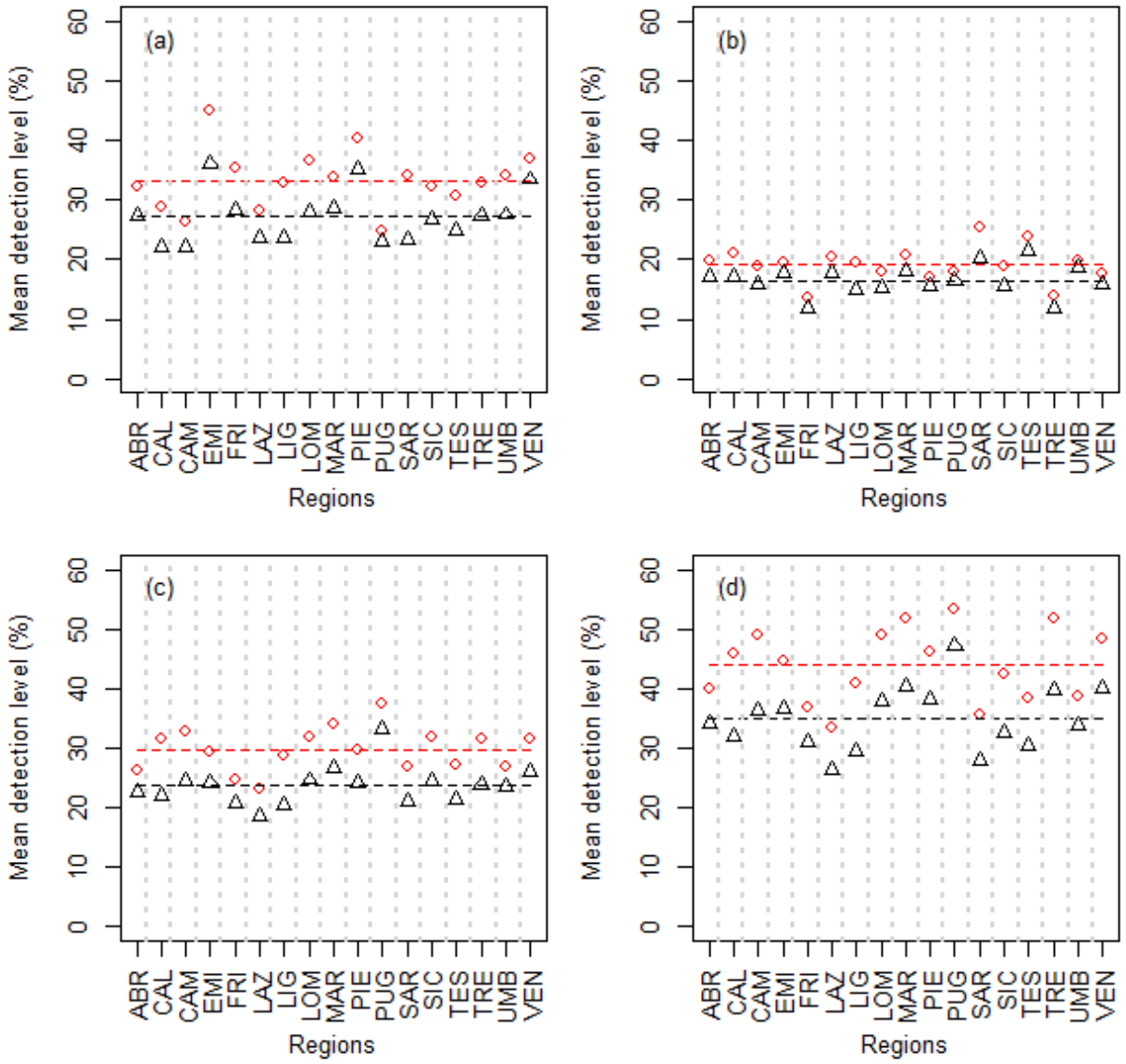

Figure 3. Overall mean detection level for 17 Italian regions from 2000 to 2010 (a) with the addition of beet sugar, (b) cane sugar, (c) water ($\delta^{18}O$ = −9‰) and (d) water ($\delta^{18}O$ = −5‰). Dashed horizontal lines correspond to overall mean detection level for the (in red) univariate and (in black) multivariate methods. Standard abbreviations for the Italian regions are as follows: ABR, Abruzzo;CAL, Calabria; CAM, Campania; EMI; Emilia-Romagna; FRI, Friuli-Venezia Giulia; LAZ, Lazio; LIG, Liguria; LOM, Lombardy; MAR, Marche; PIE, Piedmont; PUG, Puglia; SAR, Sardinia; SIC, Sicily; TES, Tuscany; TRE, Trentino-Alto Adige, UMB, Umbria; VEN, Veneto.

It should be noted that in adulteration assessments other considerations (geography, climate, year and day of harvest, and cultivar) play an important role besides the statistical evaluation and that the detection level as reported here only relates to the latter. Moreover,

for wines with precisely defined origins, like protected designation of origin wines, the statistical approach alone will have much more power, and detection levels will be lower.
Likewise, Figure 4 shows the detection rate for the different adulterations. In all cases, the dashed line indicating the multivariate approach is above the dotted line of the univariate approach. Again, this shows that the multivariate approach is more effective in detecting adulteration. Focusing for example on samples with the addition of up to 25% beet sugar, 36% of the fraud samples are detected by the univariate method and 73% with the Mahalanobis distance. Although the general trend is the same for all four adulteration types, it can be seen that the numbers are different. Adulteration with cane sugar for example is easier to detect because two isotopic ratios change rather than one. This leads to a higher detection rate for both the univariate and multivariate methods. Nevertheless, the dominance of the multivariate method over the univariate method is constant in every single case. Similarly, the results for simulations of dilution with water show that it is easier to detect adulteration with water far away from the normal range of water in wine, completely in line with expectations. Because we use two extreme values for Italy, real adulteration will be between these two values. It should be noted that in other countries, different values for the simulations are likely to be more appropriate. This does not, however, influence our main conclusion, which is that the multivariate method is much more effective in detecting adulteration than the univariate method.
It should be noted that the natural range of isotopic ratios of wine, in particular $\delta^{18}O$, varies according to geographical origin; thus, the final effectiveness of fraud detection is related to the origin of the wine and the water used for dilution. Adulteration of samples that are close to the border of the class is much easier to detect than adulteration of samples in the middle of the ellipses in Figure 1. Likewise, regions with homogeneous climatic conditions will have a much narrower CI, and in these cases it is also much easier to detect adulteration. Furthermore, values for detection level correspond to the proportion of added sugar (cane or beet) or water in the total sugar or water fraction. For juices with low natural grape sugar content due to unfavourable climatic conditions, such as rainy periods during grape development, any addition of sugar will immediately lead to high adulteration fractions. This means that a detection level of 50%, for example even if it seems high, can still be relevant: these climatic conditions are exactly the reason why some producers add sugar to the wine must. In other words, in this case addition of a low absolute amount of sugar can still be detected.

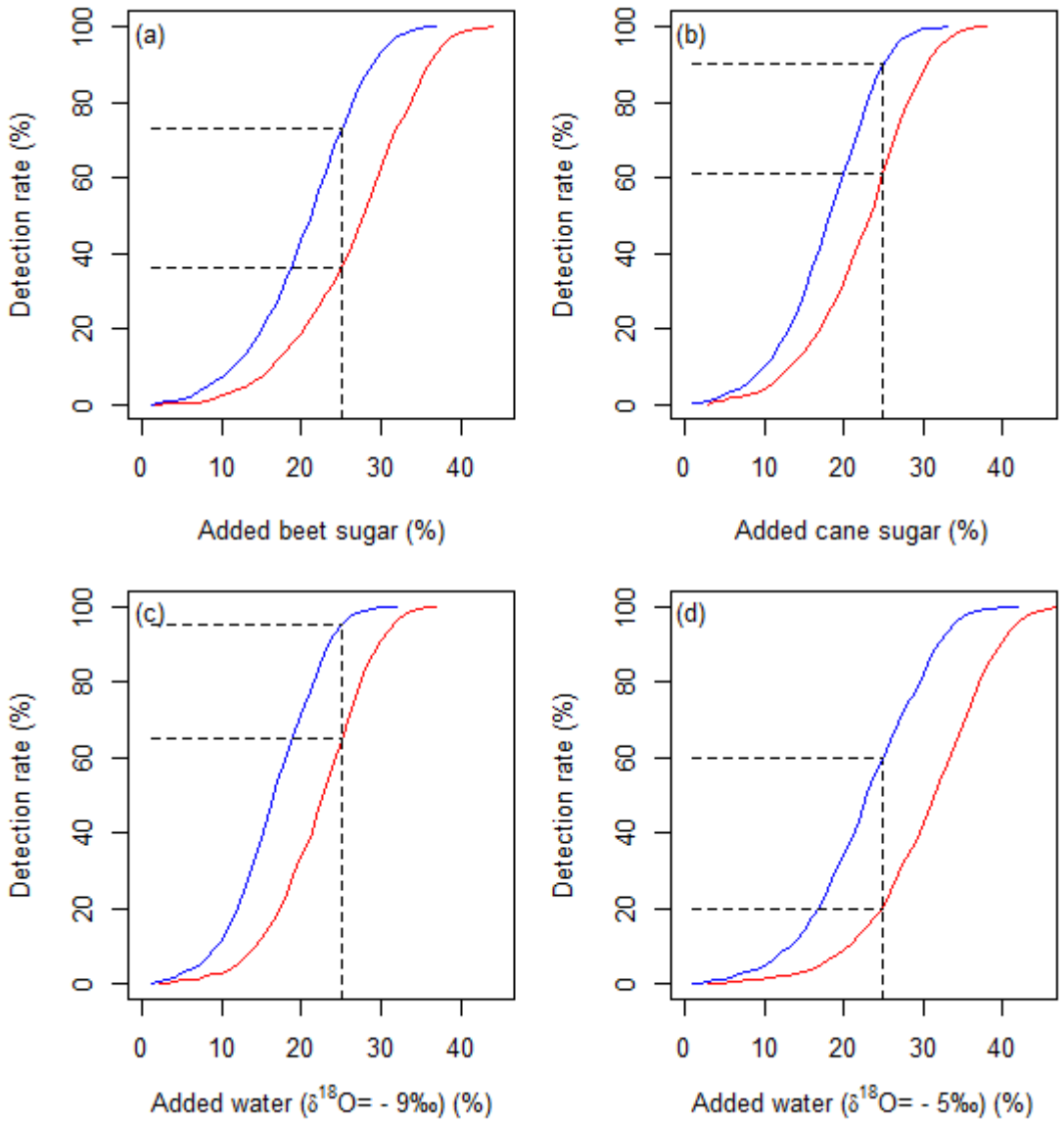

Figure 4. Detection rate for the univariate (in blue) and multivariate (in red) methods with various levels of adulteration: (a) addition of beet sugar, (b) addition of cane sugar, (c) addition of water ($\delta^{18}O = -9$) and (d) addition of water ($\delta^{18}O = -5$). Horizontal dashed lines correspond to detection rates at 25% adulteration. Tuscany 2007 samples were used for sample simulation and to determine confidence intervals.

3.4 Conclusions

This work evaluates a univariate method based on *t*-tests, currently adopted for wine authentication, using stable isotope ratios to detect the addition of sugar and water, as well as a multivariate extension based on the Mahalanobis distance. As can be seen from the results, the effectiveness of fraud detection is different. It depends on the isotopic values of

the particular sample concerned, its distance to the class centre and the type of adulteration. For some samples, even a small addition of sugar or water can be detected efficiently, while for others large amounts of sugar and water can be added before being detected. By using the wealth of information available in the European Wine Databank and by simulating realistic samples and adulterations, it is clear that significant improvements are obtained using multivariate methods. They are more effective, i.e. they detect more adulterations at a given level and are able to detect adulteration at a lower level. These results were obtained by assuming multivariate normal distribution for wine samples of one particular class (combination of region and vintage). Additional improvements would be possible using an approach based on mixture modelling for cases in which the data are not normally distributed. Although in the majority of cases in the wine databank normal distribution would seem adequate, there are cases in which a combination of two multivariate normal distributions would lead to better description of the data (Dordevic et al. 2012). For the main purpose of this paper, this is less important – the difference between multivariate and univariate detection is much greater than the difference between normally and nonnormally distributed data.

References

Capron, X., Smeyers-Verbeke, J. and Massart, D.L. (2007) Multivariate determination of the geographical origin of wines from four different countries. Food Chemistry **101**, 1585–1597.

Christoph, N., Rossmann, A. and Voerkelius, S. (2003) Possibilities and limitations of wine authentication using stable isotope and meteorological data, data banks and statistical tests. Part 1: wines from Franconia and Lake Constance 1992 to 2001. Mitteilungen Klosterneuburg **53**, 23–40.

De Maesschalck, R., Jouan-Rimbaud, D. and Massart, D.L. (2000) The Mahalanobis distance. Chemometrics and Intelligent Laboratory Systems **50**, 1–18.

Dordevic, N., Wehrens, R., Postma, G.J., Buydens, L.M.C. and Camin, F. (2012) Statistical methods for improving verification of claims of origin for Italian wines based on stable isotope ratios. Analytica Chimica Acta **757**, 19–25.

European Commission (1990) Commission regulation (EEC) No 2676/90 of 17 September 1990 determining Community methods for the analysis of wines. Official Journal of the European Union L **272**, 1–192.

European Commission (1991a) Commission Regulation (EEC) No 2347/91of 29 July 1991 on the collection of samples of wine products for the purposes of cooperation between Member States and for analysis by nuclear magnetic resonance, including analysis for the purposes of the Community databank. Official Journal of the European Union L **214**, 32–38.

European Commission (1991b) Commission Regulation (EEC) No 2348/91 of 29 July 1991 establishing a databank for the results of analyses of wine products by nuclear magnetic resonance of deuterium. Official Journal of the European Union L **214**, 39–43.

European Commission (1997) Commission Regulation (EC) No 822/97 of 6 May 1997 amending Regulation (EEC) No 2676/90 determining Community methods for the analysis of wines. Official Journal of the European Union L **117**, 10–12.

European Commission (2003) Commission Regulation (EC) No. 440/2003 of 10 March 2003 amending Regulation (EEC) No 2676/90 determining Community methods for the analysis of wines. Official Journal of the European Union L **66**, 15–23.

European Commission (2007) Commission Regulation (EC) No. 1234/2007 of 22 October 2007 establishing a common organisation of agricultural markets and on specific provisions for certain agricultural products (Single CMO Regulation). Official Journal of the European Union L **299**, 1–149.

European Commission (2008a) Commission Regulation (EC) No. 479/2008 of 29 April 2008 on the common organisation of the market in wine, amending Regulations (EC) No 1493/1999, (EC) No 1782/2003, (EC) No 1290/2005, (EC) No 3/2008 and repealing Regulations (EEC) No 2392/86 and (EC) No 1493/1999. Official Journal of the European Union L **148**, 1–61.

European Commission (2008b) Commission Regulation No. 555/2008 of 27 June 2008 laying down detailed rules for implementing Council Regulation (EC) No 479/2008 on the common organisation of the market in wine as regards support programmes, trade with third countries, production potential and on controls in the wine sector. Official Journal of the European Union L **170**, 1–80.

Gremaud, G., Quaile, S., Piantini, U., Pfammatter, E. and Corvi, C. (2004) Characterization of Swiss vineyards using isotopic data in combination with trace elements and classical parameters. European Food Research and Technology **219**, 97–104.

Holmberg, L. (2010) Wine fraud. International Journal of Wine Research **2**, 105–113.

Jolliffe, I.T. (2002) Principal component analysis (Springer series in statistics) (Springer: New York, USA).

Longinelli, A. and Selmo, E. (2003) Isotopic composition of precipitation in Italy: a first overall map. Journal of Hydrology **270**, 75–88.

Martin, G.J., Guillou, C., Martin, M.L., Cabanis, M.T., Tep, Y. and Aerny, J. (1988) Natural factors of isotope fractionation and the characterization of wines. Journal of Agricultural and Food Chemistry **36**, 316–322.

Organisation Internationale de la Vigne et du Vin (2001) Resolution Oeno 17-2001. Détermination du rapport isotopique du l'éthanol (Office Internationale de la Vigne et du Vin: Paris, France).

Organisation Internationale de la Vigne et du Vin (2009) Resolution Oeno 353-2009. Méthode de détermination du rapport isotopique 180/160 de l'eau dans le vin et le moût (Office Internationale de la Vigne et du Vin: Paris, France).

Organisation Internationale de la Vigne et du Vin (2011) Resolution Oeno 426-2011. Détermination de la repartition du deuterium dans l'éthanol issu de moûts de raisins fermentés, moûts de raisins concentres, moûts de raisins concentres rectifies etde vins par resonance magnétique nucléaire (SNIFNMR/RMNFINS) (Office Internationale de la Vigne et du Vin: Paris, France).

Ogrinc, N., Košir, I.J., Kocjančić, M. and Kidrič, J. (2001) Determination of authenticity, regional origin, and vintage of Slovenian wines using a combination of IRMS and SNIF-NMR analyses. Journal of Agricultural and Food Chemistry **49**, 1432–1440.

Roßmann, A., Schmidt, H.L., Reniero, F., Versini, G., Moussa, I. and Merle, M.H. (1996) Stable carbon isotope content in ethanol of EC data bank wines from Italy, France, and Germany. Zeitschrift für Lebensmittel- Untersuchung und -Forschung A **203**, 293–301.

Van der Veer, G., Voerkelius, S., Lorentz, G., Heiss, G. and Hoogewerff, J.A. (2009) Spatial interpolation of the deuterium and oxygen-18 composition of global precipitation using temperature as ancillary variable. Journal of Geochemical Exploration **101**, 175–184.

Van Jaarsveld, F.P. (2008) Authenticity of South African wines (WW 08/26). http://www.sawislibrary.co.za/dbtextimages/finalreport148.pdf

Varmuza, K. and Filzmoser, P. (2009) Introduction to multivariate statistical aanalysis in chemometrics (Taylor & Francis – CRC Press: Boca Raton, FL, USA).

Wachter, H., Christoph, N. and Seifert, S. (2009) Verifying authenticity of wine by Mahalanobis distance and hypothesis testing of stable isotope pattern – a case study using the EU Wine Databank. Mitteilungen Klosterneuburg **59**, 237–249.

Chapter 4

4 Climatic and geographical dependence of the H, C and O stable isotope ratios of Italian wine

F. Camin **, N. Dordevic**, R. Wehrens, M. Neteler, L. Delucchi, G.J. Postma, L. M. C. Buydens
Analytica Chimica Acta (2015), 853, pp 384-390.

Abstract

In this study, we investigated the relationship between $(D/H)_1$, $(D/H)_2$ and $\delta^{13}C$ of ethanol and $\delta^{18}O$ of water in wine, and variables describing the climate and the geography of the production area, using exploratory visualisation tools, regression analysis and linear modelling. For the first time, a large amount of data (around 4000 wine samples collected over 11 years in Italy) and all the official isotopic parameters, as well as a large number of significant climatic and geographical descriptors (date of harvest, latitude, longitude, elevation, distance from the sea, amount of precipitation, maximum daily temperature, minimum daily temperature, mean daily temperature, $\delta^{18}O$ and $\delta^{2}H$ of precipitation) were considered. $\delta^{18}O$, followed by $(D/H)_1$, was shown to have the strongest relationship with climate and location. The dominant variables were latitude, with a negative relationship, $\delta^{18}O$ and $\delta^{2}H$ of precipitation and temperature, both with positive relationships. The identified correlations and models could be used to predict the isotopic composition of authentic wines, offering increased possibilities for detecting fraud and mislabelling.

Keywords

Wine; Isotope ratio mass spectrometry; Climate; Geography; Prediction; Mislabelling

4.1 Introduction

The site-specific D/H isotope ratio in the methylic and methilenic site of ethanol, $(D/H)_1$ and $(D/H)_2$, the $^{13}C/^{12}C$ ($\delta^{13}C$) ratio of ethanol and the $^{18}O/^{16}O$ ($\delta^{18}O$) ratio of water have been analysed in wine since 1987 [1] to determine fraud, such as sugar addition or chaptalisation, watering down and mislabelling, i.e. false declarations of origin. The variability of these isotopic ratios is indeed related to the botanical origin of the sugar (grape, cane or beet) and to the climatic and geographical characteristics of the area where the plant grows. The relationship between the isotopic data of wine and physical variables describing the climate and geography of the production area is an interesting topic, as reflected in the high number of related papers published in the last 20 years (see background). However, only a few of these studies considered a large number of samples and more than one isotopic ratio, and none took into account a high number of physical variables which could influence the isotopic values of wine.

In this study, we considered around 4000 Italian wine samples produced all over Italy in 11 consecutive years from 2000, all the official isotopic parameters and a large number of potentially influential factors reflecting the geography and climate of the production locations. Exploratory visualisation tools, regression analysis and linear modelling were used in our study to investigate the most significant relationships and define multivariate models able to infer isotopic values from climatic and geographical data. These correlations and models, once validated in further studies, could be used to predict wine isotopic data that could be adopted to define authenticity confidence limits to refer to in authenticity verification.

4.1.1 Background

In previous studies, $(D/H)_1$, $(D/H)_2$ and $\delta^{13}C$ of ethanol were found to be positively correlated with mean temperature and negatively correlated with the amount of precipitation during the sugar accumulation and grape ripening period in wines from 34 regions across the world [2], as well as on a local Italian scale [3]. $\delta^{18}O$ of wine water also showed linear correlation with temperature and the amount of precipitation [2], but was also influenced by other climatic parameters. Indeed, in around 40 wines from California, good correlation was found between $\delta^{18}O$ and the average daily relative humidity in the 3 week period prior to harvest and crop evapotranspiration in the month of September, in

addition to the average maximum daily temperatures from July [4]. By applying stepwise multiple regression analysis, West et al. [5] showed that the monthly mean maximum daily temperature and monthly mean daily dew point temperature in September and October and $\delta^{18}O$ of local precipitation were effective in predicting $\delta^{18}O$ of wine water for around 50 wines from the western coast of the USA, and developed a predictive model on this basis. Hermann and Voerkelius [6] developed a different model for $\delta^{18}O$, tested on 775 wines from 6 vintages in Germany, where relative humidity for a period of 30 days prior to harvest and the $\delta^{18}O$ of atmospheric humidity played a predominant role. As a consequence of climatic variation in different years, the year of production was shown to be a critical factor [6–11], along with the time of harvest, in the case of wines from regions which experience climatic instability and rapid changes in climatic parameters. Normally, early harvested wines have higher $\delta^{18}O$ values than late harvested wines [7]. Besides the climatic conditions in the production area, irrigation can affect $\delta^{18}O$ and $\delta^{13}C$ in wine, especially in a hot climate, significantly decreasing values [12]. No effect of irrigation on $\delta^{13}C$ was observed in northern Italian wines [13] and it was found to be minimal in Californian wines [4]. The grape variety was also found to affect wine isotopic values (especially $\delta^{18}O$ and $\delta^{13}C$), but this effect is normally a consequence of differences in the ripening period or harvest data [3,7,10] or can be confused by different climates in the area of origin [5,9]. Only Gomez-Alonso and García-Romero [12] observed that $\delta^{18}O$ and $\delta^{13}C$ were affected by variety, regardless of the sampling data, in a study on a local scale. As regards geographical influence on $(D/H)_1$ and $(D/H)_2$, 284 wines selected from 19 countries from outside Europe correlated with latitude, except for countries which use irrigation water originating from a higher altitude in the mountains [14]. A trend for decreasing values in all these three isotopic parameters in relation to latitude was observed in Italian wines [11] and for $\delta^{18}O$ in around 80 southern Brazilian wines, also found to be related to altitude [9]. Distance from the sea was noted to be a significant variable for 102 samples from Slovenia [8]: different D/H, $\delta^{13}C$ and $\delta^{18}O$ values were reported for coastal and continental regions, but this could also be a consequence of the different climate in the regions [15]. Finally, differences between the Adriatic and Tyrrhenian coast were found for $(D/H)_1$ of Italian wines in the period 1987–1992 [16], highlighting the significance of longitude as a factor.

4.2 Materials and methods

4.2.1 Wine isotope data

The isotope data of 3948 Italian wines produced from 2000 to 2010 (from 303 to 421 per year) were evaluated. For each sample four isotopic parameters were available. The ratio between the deuterium and hydrogen of the methyl site of wine ethanol [$(D/H)_1$] and the corresponding ratio of the methylene site [$(D/H)_2$] were determined with site-specific natural isotope fractionation– nuclear magnetic resonance (SNIF–NMR) of deuterium. The other two isotopic ratios ($^{13}C/^{12}C$ of ethanol and $^{18}O/^{16}O$ of water) were measured using isotope ratio mass spectrometry (IRMS). The isotopic composition determined using IRMS was denoted in delta [17] in relation to the international standard VPDB (Vienna Pee Dee Belemnite) for $\delta^{13}C$ and VSMOW (Vienna Standard Mean Ocean Water) normalized to the VSMOW–SLAP (Standard Light Antarctic Precipitation) scale, for $\delta^{18}O$. The SNIF–NMR data are expressed in ppm, whereas the IRMS data in ‰. All data were obtained according to the official methods (OIV methods MA-AS2-12, MA-AS312-06, MA-AS311-05 [18]). Additional information is available for each sample, such as origin (at municipal level) and the exact date of harvest.

4.2.2 Climatic/geographical data

The geographical and climatic parameters (Table 1) were obtained for each sample, based on the location of the respective vineyard (i.e. the municipality where the vineyard is situated) and the date of harvest. The data source and resolution are specified in Table 1. Latitude, longitude, elevation and distance from the sea were considered as geographical variables for each vineyard. In order to determine climatic impact, temperatures (daily average, minimum and maximum) and the amount of precipitation (mm/day) were taken into account. $\delta^{18}O$ and $\delta^{2}H$ values of precipitation, which are analogous to $\delta^{18}O$ and $\delta^{2}H$ values of soil water and thus of source of grape water, were also considered. Due to the large variations in data from different years and different harvest dates, the year and date of harvest were also taken into account. Thus, all the parameters identified as significant in the literature (see Section 1.1) were considered, with the exception of humidity, $\delta^{18}O$ of water vapour and evapotranspiration. Reliable values for relative humidity on a local scale are simply not available. Moreover, relative humidity is correlated with the level of precipitation over the long-term [19]. $\delta^{18}O$ values of water vapour are not available, as they vary between years and during the year [20] and can only be predicted [6]. The variety and irrigation were also not evaluated, because they were often identified as not significant (see Section 1.1).

The date of harvest was converted to the corresponding day of the year number (an integer between 1 and 365 or 366, depending on the number of days in the year of harvest). The positional data (latitude, longitude) were retrieved from the web, on the basis of the municipality where the vineyard is located, and were then processed using GRASS GIS [21] to extract the respective vineyard elevation and distance from sea. The climatic parameters were derived from daily gridded datasets of the European Climate Assessment & Dataset (ECA&D; [22]) using the temporal frame-work in GRASS GIS 7 [23]. The $\delta^{18}O$ and $\delta^{2}H$ values of precipitation are accessible on the internet for any point on the earth [24].

Table 1. Description of geographic, climatic and isotopic data used in this study.

Variables	**Abbreviation**	**Data type**	**Resolution**	**Source**
Date of harvest	H	static		
Latitude	y	static	point	web
Longitude	x	static	point	web
Elevation	E	static	20m	Italian elevation model
Distance from the sea	D	static	250m	Derived from elevation map in GIS
Amount of precipitation [mm/day]	P	dynamic	25km	ECA&D, http://www.ecad.eu
Maximum daily temperature	mean T	dynamic	25km	ECA&D
Minimum daily temperature	min T	dynamic	25km	ECA&D
Mean daily temperature	max T	dynamic	25km	ECA&D
$\delta^{18}O$ MW	$\delta^{18}O$ MW	static	37km	[24]
$\delta^{2}H$ MW	$\delta^{2}H$ MW	static	37km	[24]

4.2.3 Exploratory analysis

To assess the relationships between all the isotopic, climatic and geographical parameters, exploratory analysis was performed using correlation analysis and principal component analysis (PCA) [25].

4.2.4 Linear modelling

The next step was to predict the isotopic ratios in wine using geographical and climatic information. The vintage, or year, was also included, since this may capture climatic variations that are not taken into account in the precipitation and temperature variables. Since our main aim is interpretation of the models, rather than achieving optimal prediction, some highly correlated variables were eliminated from the model; we used only mean temperature from the cluster of highly correlated temperature measurements and we did not consider the isotopic values of precipitation, since they are highly correlated with geographical position (see below in the Section 3). The equation for the overall model thus becomes:

$$(1)\ SIR = b_A A + b_X X + b_Y Y + b_E E + b_T T + b_P P + b_H H + b_D D$$

where SIR is any of the four stable isotope ratios and the eight variables are year (A), longitude (X), latitude (Y), elevation (E), the average mean temperature (T) and the amount of precipitation (P) over fifteen days prior to harvest, harvest date (H) and distance from the sea (D). The corresponding coefficients are indicated with a b and a variable-dependent subscript.

A stepwise selection approach was used, starting with the full model and eliminating (and possibly reinstating) variables according to Akaike's information criterion [26,27]. We treated the year factor as a categorical variable, which basically served to allow for a constant offset in each year, the other coefficients being the same across all years.

All statistical analysis was done using the statistical software R [28].

4.3 Results and discussion

The isotope ratios of Italian wine samples show clear spatial variations. An example is shown in Figure 1, illustrating the mean values for the four isotopic ratios of Italian wines, averaged for each of the Italian regions for the 2001 vintage. It is this spatial variation that we wish to model using climatic and geographical variables.

Some of the geographical and climatic parameters (see Table 1) are static (single value), while others are dynamic (series of values on the basis of the period before harvest considered).

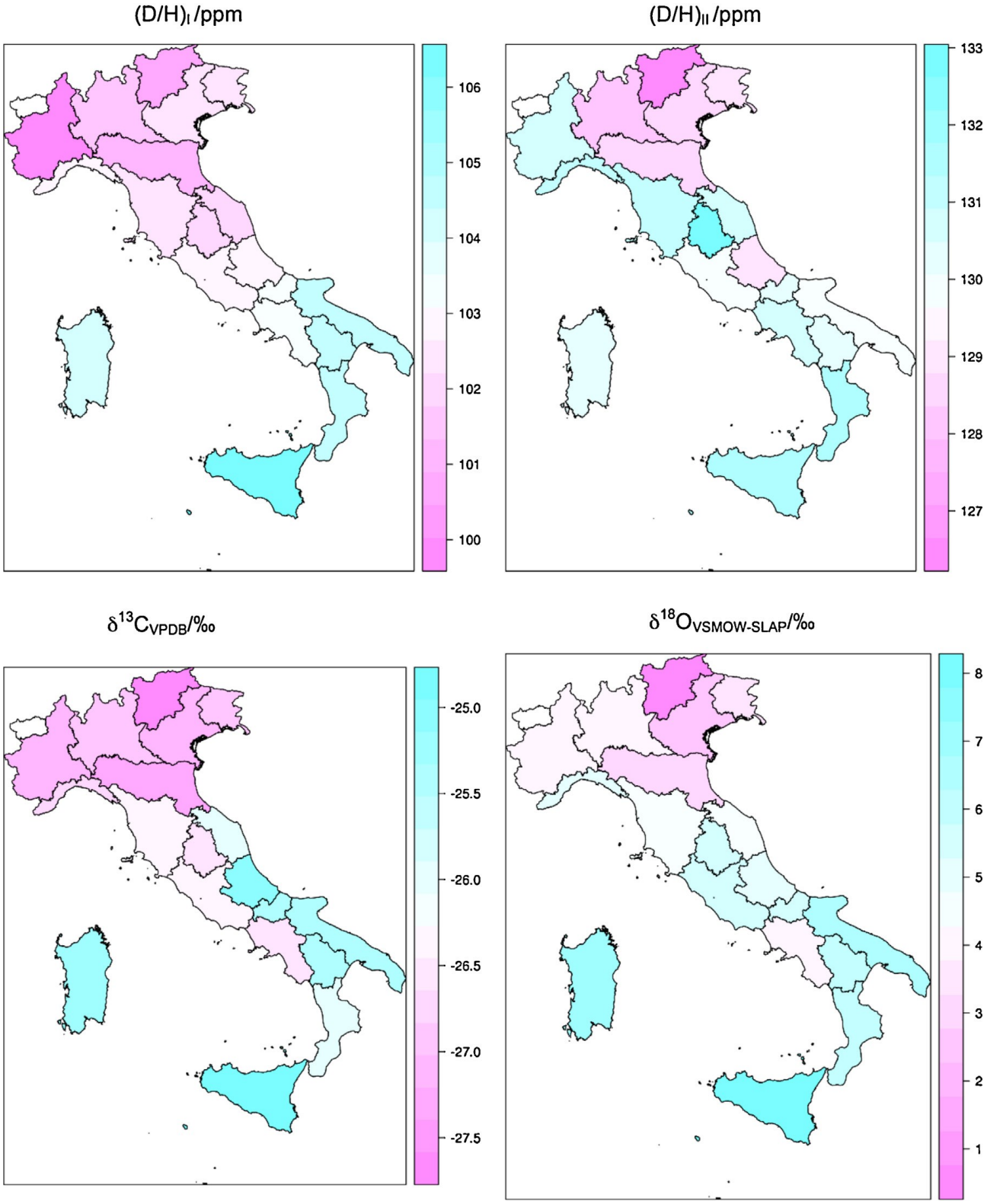

Figure 1. Regional mean values for the four isotope ratios in wine samples from Italy, 2001 (the Valle d'Aosta is white, because the values are not available).

As regards dynamic variables, it is important to define which period of time before harvest contains most information, in order to select the data for that period. In the literature there are contrasting findings. For Bigwood et al. [14], $\delta^{18}O$ of grape juice is influenced by precipitation events just prior to harvest and hence the short time period during the final development of the grape is relevant. In contrast, for D/H of sugar and then ethanol a much longer period is relevant, because sugars are biosynthetised over the entire time-period in which the grape develops. A strong correlation between wine $\delta^{18}O$ and climate immediately prior to harvest has been reported in Rossmann et al. [29] and in Christoph et al. [30] whereas others authors found stronger correlations by including the warmer midsummer months [4]. Using stepwise multiple regression analysis, West et al. [5], showed both September and October temperature to be very effective in terms of prediction. Finally, Hermann and Voerkelius [6] considered the arithmetic mean values of temperature and humidity for a period of 30 days prior to harvest, to cover the most relevant period of grape maturation.

To choose the most informative period before harvest, all the combinations of mean temperature and precipitation values over different periods (days) prior to harvest were tested using linear modelling, fitting linear regression models together with other available parameters in Table 1. We did not find any significant effects on $(D/H)_1$, $(D/H)_2$ and $\delta^{13}C$ when using different periods of temperature and precipitation before harvest. Only in the case of $\delta^{18}O$, did R^2 slightly increase from the day of harvest, reaching its maximum 2 weeks before harvest. There was no significant improvement in the R^2 value by taking into account longer periods of up to 60 days. Thus the mean values for temperature and amount of precipitation over a period of 15 days before harvest were taken into account for statistical analysis.

4.3.1 Exploratory analysis

The correlation matrix between all variables, except the year of harvest (which is treated as a discrete factor) is presented in Figure 2. There is a positive correlation between all wine isotopic parameters and high correlations between mean, minimum and maximum temperatures. Moreover, there is a very high linear correlation (around 1) between $\delta^{18}O$ and $\delta^{2}H$ of precipitation water, as described in the literature [31] and a high negative correlation between $\delta^{18}O$ and $\delta^{2}H$ of precipitation water and latitude (-0.85 and -0.88).

As regards isotopic variables of wine, $\delta^{18}O$ and $(D/H)_1$ showed a higher correlation with climatic and geographical parameters (up to 0.5). $\delta^{18}O$ was mainly positively related with

the isotopic values of water and with temperatures, and negatively correlated with precipitation, date of harvest, latitude and distance from the sea. $(D/H)_1$ was mainly positively related with the isotopic values of water and negatively correlated with latitude and distance from the sea.

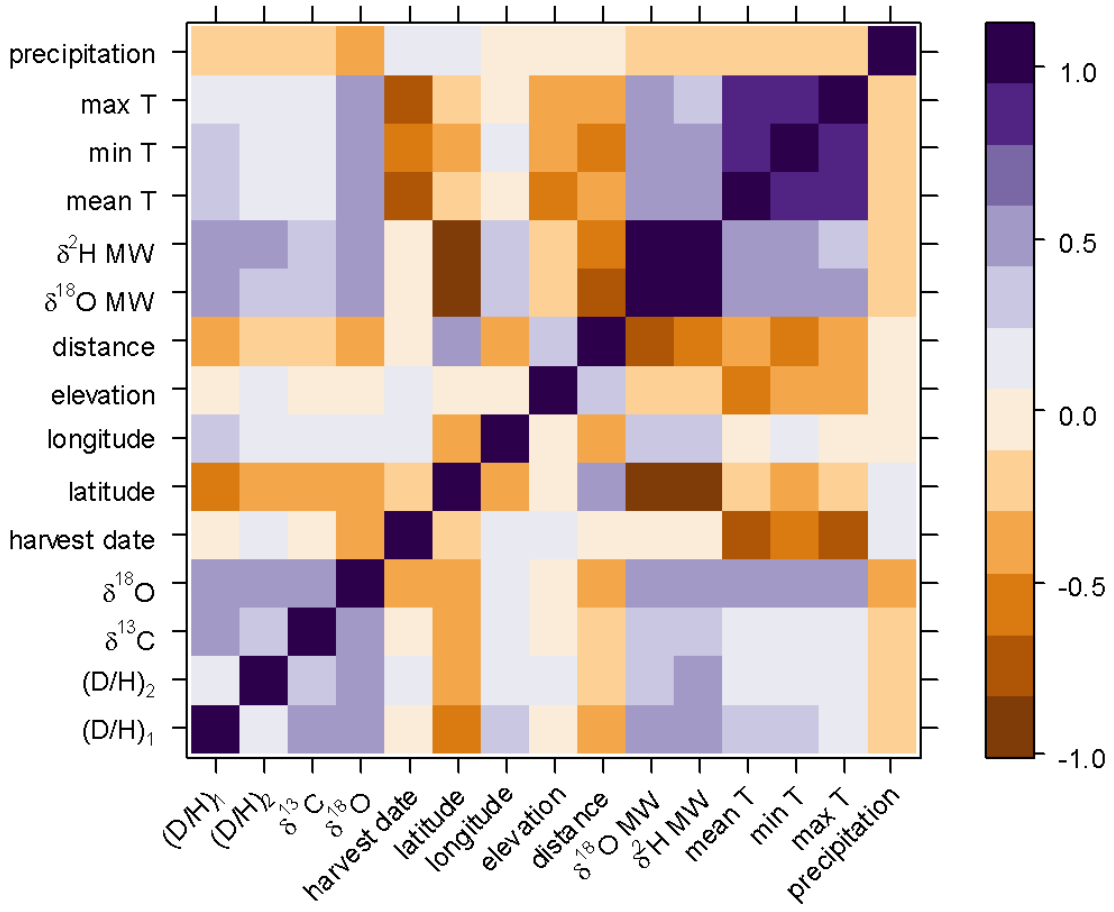

Figure 2. Correlation matrix between $(D/H)_1$, $(D/H)_2$, $\delta^{13}C$ and $\delta^{18}O$ parameters of wine and the climatic and geographical characteristics of the areas of origin.

Similar information can be obtained from principal component analysis. The plot in Figure 3 shows the loadings of all variables in the space of the first two PCs, with 58% of total variance explained. Again we see reasonably high correlations between the four isotopic ratios in the wine on the one hand, and these isotopic ratios and the isotopic ratios in precipitation on the other (the arrows point in more or less the same direction). The mean, minimum and maximum temperatures averaged over a period of 15 days before harvest were highly correlated. Higher temperatures were anticorrelated with the harvest date,

which indicates that in warmer areas the harvest date occurs earlier in the year. Similarly, higher elevation was coupled with lower temperatures and later harvest dates. The four isotopic ratios in the wine show reasonable correlations with latitude and longitude, confirming the geographical influence on wine isotope ratios also shown in Figure 1.

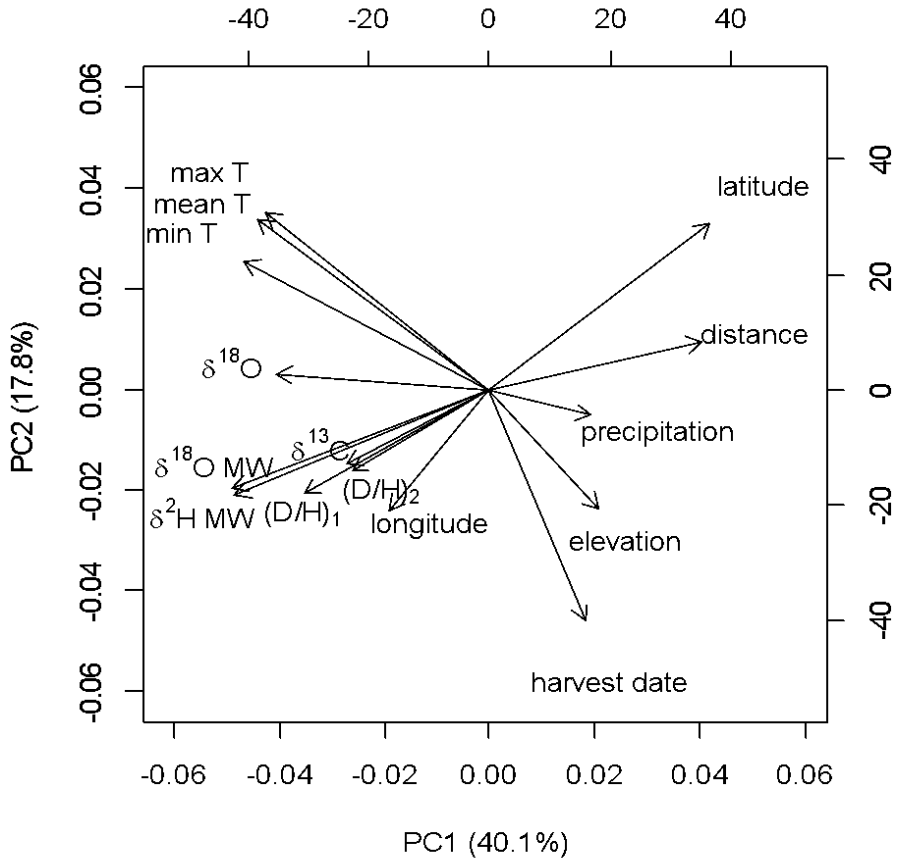

Figure 3. Loading plot for wine samples from 2000 to 2010, together with corresponding isotopic, climatic and geographical information.

4.3.2 Linear modelling

Fitting linear models to the four isotopic ratios in every case led to significant models, clearly showing that isotopic ratios are influenced by climatic and geographical parameters. This is most obvious for the prediction model of $\delta^{18}O$ (R^2 = 0.71), indicating that this is a strong potential marker for wine origin determination. The second most significant marker is $(D/H)_1$ (R^2 = 0.42). $(D/H)_2$ and $\delta^{13}C$ (R^2 = 0.30 and 0.24, respectively) had less clear relationships with climate and location.

Values of the coefficients for individual years, b_A, are shown in Figure 4. In each case, the first year considered here, 2000, is taken as a reference value (0). For $(D/H)_1$, it is clear that in all years except 2003 the average values were lower than in 2000. Something similar is observed in the case of $\delta^{18}O$; it is a consequence of the particularly hot Italian climate in the pre-harvest period in 2000 and 2003 and the cold and rainy climate in 2002, 2005 and 2010.

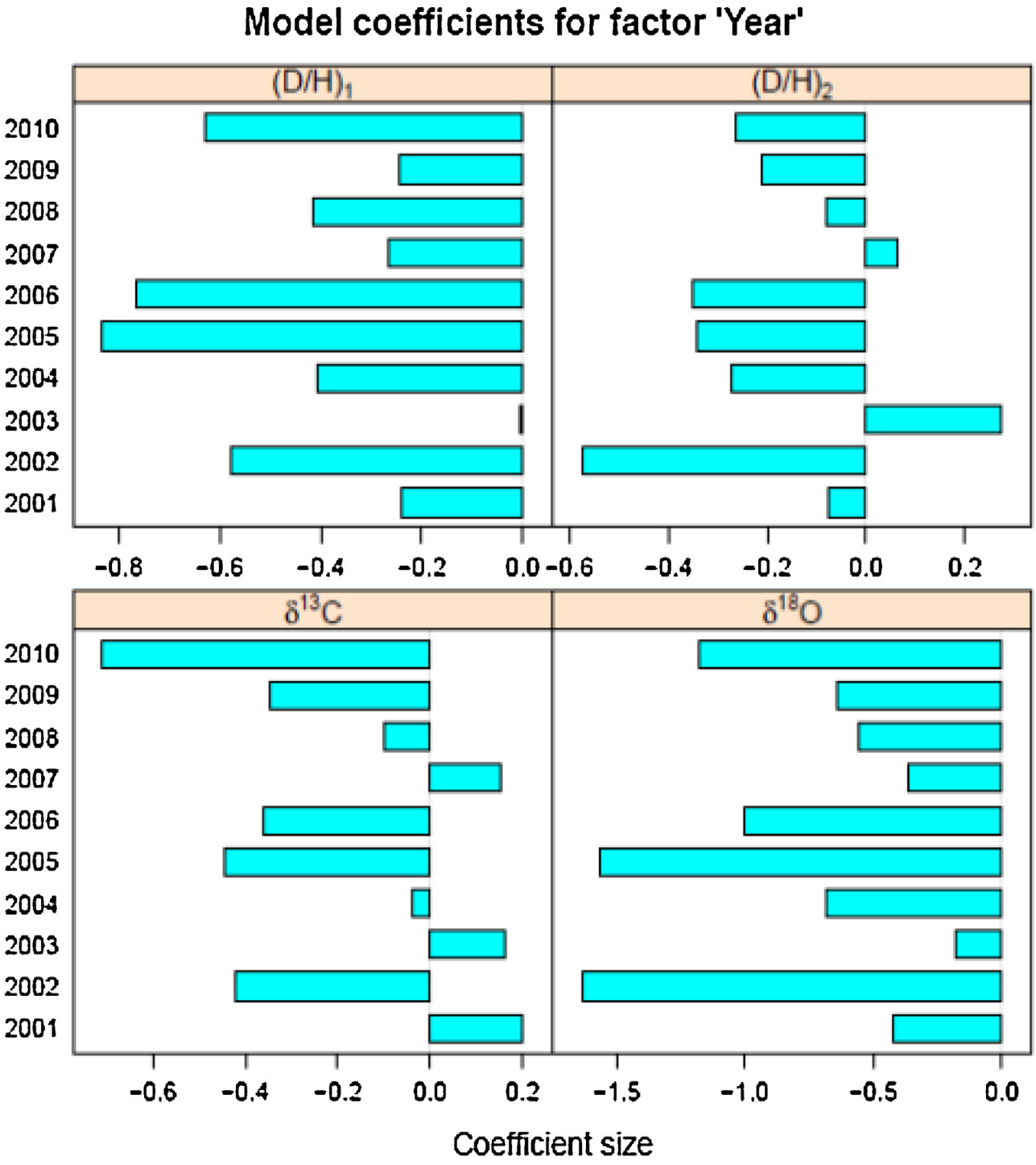

Figure 4. Coefficients for the factor 'year' in the linear model (1). 2000 is taken as a reference.

The coefficients for the other continuous variables are summarised in Table 2. To make the values comparable, prior to linear modelling, we standardised all the variables to zero mean and unit variance (note that this does not influence the linear models since they are affine equivariant). In the case of $(D/H)_2$ and $\delta^{18}O$, all seven variables were included in the final model, although the coefficient for elevation is not significant for $\delta^{18}O$. For the other two isotopic ratios, some variables were removed from the stepwise procedure.

Table 2 clearly shows that the dominant variable is the North– South coordinate, namely latitude. For all four ratios, lower values were found for higher latitudes. The significance of latitude has been already observed in Italian wines [11], but also in non- European wines [9,14]. Other variables with similar effects for all isotope ratios are the distance from the sea (a greater distance from the sea is associated with lower isotope ratios, except for $\delta^{13}C$ where no effect is observed), as also observed in Slovenian wines [8], and temperature (higher temperatures correspond to higher isotope ratios, again with the exception of $\delta^{13}C$). The last observation is in agreement with the literature for wines with a different geographical origin produced all over the world [2,4–6].

Table 2. Regression coefficients for the continuous variables in Eq. (1). Numbers in bold indicate significant coefficients at $p < 0.05$ level. Empty cells indicate that stepwise selection did not include this variable in the final equation.

	Longitude (x) b_X	**Latitude (y) b_Y**	**Elevation b_E**	**Distance to the sea b_D**	**Temperature b_T**	**Precipitation b_P**	**Date of harvest b_H**
$(D/H)_1$	**0.065**	**-0.459**	**-0.088**	**-0.097**	**0.061**		**0.071**
$(D/H)_2$	**-0.169**	**-0.387**	**0.101**	**-0.057**	**0.220**	**-0.031**	**0.225**
$\delta^{13}C$	**-0.080**	**-0.410**			0.003	-0.020	**-0.061**
$\delta^{18}O$	**-0.141**	**-0.447**	-0.011	**-0.034**	**0.145**	**-0.136**	**-0.185**

$(DH)_1$ and $(D/H)_2$ increased with the date of harvest, whereas $\delta^{13}C$ and $\delta^{18}O$ decreased. In the literature only $\delta^{18}O$ was found to decrease with the date of harvest [7]. $(D/H)_1$ increased with longitude, as observed for Italian wines from the period 1987–1992 [16], whereas $(D/H)_2$, $\delta^{13}C$ and $\delta^{18}O$ decreased. Differences between the Adriatic and Tyrrhenian coasts were also observed in $\delta^{2}H$ of bulk Italian olive oils, as a consequence of the different source and isotopic composition of rainfall and the different climatic conditions

on the two coasts [32]. $(D/H)_1$ and $\delta^{18}O$ decreased with elevation (although not at the 0.05 significance level), as observed in Brazilian wines [9], whereas $(D/H)_2$ tended to increase.

4.4 Conclusions

This is the first comprehensive work investigating the relation between wine isotopic ratios and environmental variables, based on a large number of samples collected over 11 years from a large and climatically diverse area such as Italy. Wine $(D/H)_1$, $(D/H)_2$, $\delta^{13}C$ and $\delta^{18}O$ were evaluated in relation to a number of climatic and geographical descriptors of their areas of origin. In previous works in the literature to date, these phenomena have only been evaluated on the basis of a few variables and with a limited scope. Relationships between climatic and geographical parameters were evaluated here, providing a broader insight into these relations. $\delta^{18}O$, followed by $(D/H)_1$, showed the strongest relationship with climate and location. Considering the results of both exploratory analysis and linear modelling, we can conclude that the dominant variables are latitude, with a negative relationship, and $\delta^{18}O$ and $\delta^{2}H$ of precipitation and temperature, both with positive relationships. Although correlation-based models such as the ones proposed here cannot be used to deduce causation, we did obtain additional information. Some variables are likely to have causal effects on wine stable isotope ratios (temperature, amount of precipitation, $\delta^{2}H$ and $\delta^{18}O$ of precipitation) whereas others (e.g. latitude) are more likely to be proxies for other, unknown variables, or variables inaccessible for measurement on a country-wide scale.

These relationships may be used to predict the isotopic composition of wines, using appropriate climatic characteristics for the production region. This may open up possibilities for improved protection of both users and producers, offering increased possibilities for detecting fraud and mislabelling.

Acknowledgments

We acknowledge the Central Anti-fraud Department of the Italian Ministry of Agriculture, Food and Forestry Policy in Rome, the E-OBS dataset from the EU-FP6 project ENSEMBLES (http:// ensembles-eu.metoffice.com) and the data providers in the ECA&D project (http://www.ecad.eu).

References

[1] G. Martin, S. Brun, Applications de la risonance magnetique nucleaire du deuterium au controle des mouts, des mouts concentres, du sucre de raisin et des vins (R.M.N.-F.I.N.S./S.N.I.F.-N.M.R.), Bulletin de l'OIV 671–672 (1987) 131–145.

[2] G.J. Martin, M.L. Martin, Climatic significance of isotope ratios, Phytochem. Rev. 2 (2003) 179–190.

[3] C. Aghemo, A. Albertino, R. Gobetto, F. Spanna, Correlation between isotopic and meteorological parameters in Italian wines: a local-scale approach, J. Sci. Food Agric. 1 (2011) 2088–2094.

[4] N. Ingraham, E. Caldwell, Influence of weather on the stable isotopic ratios of wines: tools for weather/climate reconstruction, J. Geophys. Res. 104 (1999) 2185–2194.

[5] J.B. West, J.R. Ehleringer, T.E. Cerling, Geography and vintage predicted by a novel GIS model of wine d18O, J. Agric. Food Chem. 55 (2007) 7075–7083.

[6] A. Hermann, S. Voerkelius, Meteorological impact on oxygen isotope ratios of German wines, Am. J. Enol. Viticult. 59 (2008) 194–199.

[7] A. Rossmann, F. Reniero, I. Moussa, H.L. Schmidt, G. Versini, M.H. Merle, Stable oxygen isotope content of water of EU data-bank wines from Italy, France and Germany, Z. Lebensm. Unters. Forsch. A 208 (1999) 400–407.

[8] N. Ogrinc, I.J. Kosir, M. Kocjancic, J. Kidric, Determination of authenticity, regional origin, and vintage of Slovenian wines using a combination of IRMS and SNIF–NMR analyses, J. Agric. Food Chem. 49 (2001) 1432–1440.

[9] L. Adami, S.V. Dutra, Â.R. Marcon, G.J. Carnieli, C.A. Roani, R. Vanderlinde, Geographic origin of southern Brazilian wines by carbon and oxygen isotope analyses, Rapid Commun. Mass Spectrom. 24 (2010) 2943–2948.

[10] D.A. Magdas, S. Cuna, G. Cristea, R.E. Ionete, D. Costinel, Stable isotopes determination in some Romanian wines, Isot. Environ. Health Stud. 48 (2012) 345–353.

[11] N. Dordevic, R. Wehrens, G.J. Postma, L.M.C. Buydens, F. Camin, Statistical methods for improving verification of claims of origin for Italian wines based on stable isotope ratios, Anal. Chim. Acta 757 (2012) 19–25.

[12] S. Gómez-Alonso, E. García-Romero, Effect of irrigation and variety on oxygen ($\delta^{18}O$) and carbon ($\delta^{13}C$) stable isotope composition of grapes cultivated in a warm climate, Aust. J. Grape Wine Res. 16 (2010) 283–289.

[13] S. Poni, F. Bernizzonia, S. Civardia, M. Gattia, D. Porro, F. Camin, Performance and water-use efficiency (single-leaf vs. whole-canopy) of well-watered and half-stressed split-root Lambrusco grapevines grown in Po Valley (Italy), Agric. Ecosyst. Environ. 129 (2009) 7–106.

[14] T. Bigwood, M. Sharman, A. Aldus, M.J. Dennis, Global variations in the deuterium/hydrogen isotope ratios of wine, J. Wine Res. 9 (1998) 155–166.

[15] I.J. Kosir, M. Kocjancic, N. Ogrinc, J. Kidric, Use of SNIF–NMR and IRMS in combination with chemometric methods for the determination of chaptalization and geographical origin of wines (the example of Slovenian wines), Anal. Chim. Acta 429 (2001) 197–208.

[16] A. Monetti, F. Reniero, G. Versini, Interregional and interannual variability of ethanol site specific deuterium content in Italian wines, Z. Lebensm. Unters. Forsch. 199 (1994) 311–316.

[17] T.B. Coplen, Guidelines and recommended terms for expression of stable-isotope-ratio and gas-ratio measurement results, Rapid Commun. Mass Spectrom. 25 (2011) 2538–2560.

[18] Compendium of International Methods of Analysis of Wines and Musts Vol. 1 and 2, International Organisation of Vine and Wine (OIV), 2014 http://www.oiv.int/oiv/info/enmethodesinternationalesvin

[19] H. Masunaga, Short-term versus climatological relationship between precipitation and tropospheric humidity, J. Climate 25 (2012) 7983–7990.

[20] H. Jacob, C. Sonntag, An 8-year record of the seasonal variation of ^{2}H and ^{18}O in atmospheric water vapour and precipitation at Heidelberg, Germany, Tellus B 43 (1991) 291–300.

[21] M. Neteler, M.H. Bowman, M. Landa, M. Metz, GRASS GIS: a multi-purpose open source GIS, Environ. Model. Softw. 31 (2012) 124–130.

[22] M.R. Haylock, N. Hofstra, A.M.G.K. Tank, E.J. Klok, P.D. Jones, M.A. New, European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, J. Geophys. Res. Atmos. 113 (2008) D20119.

[23] S. Gebbert, E. Pebesma, TGRASS: a temporal GIS for field based environmental modelling, Environ. Model. Softw. 53 (2014) 1–12.

[24] G.J. Bowen, J. Revenaugh, Interpolating the isotopic composition of modern meteoric precipitation, Water Resour. Res. 39 (2003) 1299–1312. http://wateriso.utah.edu/waterisotopes/pages/information/mapping.html.

[25] I.T. Jolliffe, Principal Component Analysis (Springer Series in Statistics), Springer, New York, 2002.

[26] T.J. Hastie, D. Pregibon, Generalized linear models, in: J.M. Chambers, T.J. Hastie (Eds.), Statistical Models in S, Wadsworth & Brooks/Cole, 1992.

[27] W.N. Venables, B.D. Ripley, Modern Applied Statistics with S, 4th ed., Springer, New York, 2002.

[28] R Development Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2014. . ISBN 3-900051-07-0 http://www.R-project.org/.

[29] A. Rossmann, F. Reniero, I. Moussa, H.-L. Schmidt, G. Versini, M.H. Merle, Stable oxygen isotope content of water of EU data-bank wines from Italy, France and Germany, Z. Lebensm. Unters. Forsch. A 208 (1999) 400–407.

[30] N. Christoph, A. Rossmann, S. Voerkelius, Possibilities and limitations of wine authentication using stable isotope and meteorological data, data banks and statistical tests. Part 1: wines from Franconia and Lake Constance 1992 to 2001, Mitteilungen Klosterneuburg 53 (2003) 23–40.

[31] H. Craig, Isotopic variations in meteoric waters, Science 3465 (1961) 1702–1703.

[32] L. Bontempo, F. Camin, R. Larcher, G. Nicolini, M. Perini, A. Rossmann, Coast and year effect on H, O and C stable isotope ratios of Tyrrhenian and Adriatic italian olive oils, Rapid Commun. Mass Spectrom. 23 (2009) 1043–1048.

Chapter 5

5 Synthesis and perspectives

5.1 Introduction

In this thesis we analyse the performance of methods based on the stable isotope ratios of wines for authenticity applications, using newly developed statistical procedures for additional improvement. All statistical procedures were evaluated in the real context of wine authentication, with the scope of detecting false claims of origin and the use of unpermitted substances during the production process, such as the addition of sugar and water. We also investigated to what extent the stable isotope ratios of wine are related to the climatic and geographical characteristics of grape origin. Understanding these relationships can help us to develop models able to predict isotopic composition based on the climatic and geographical characteristics of wines, and to verify the declared origin of wine.

Fraud in the wine sector is common and has taken place since ancient times. The types of fraud have changed over time and all of them are designed to mislead consumers. Wine prices on the market can vary extensively and prices depend on consumer perception and wine quality, which is affected by grape origin and the year of vintage. These are the reasons leading dishonest producers to claim a false origin for wines, thus increasing their profit. Furthermore, other unpermitted techniques are used during the winemaking process, such as the addition of sugar and water.

Due to the large number of cases and the opportunities for fraud in the wine sector, many research groups have been working on the development of reliable analytical techniques able to detect such fraud. Some of these techniques are very efficient in detecting only particular kinds of fraud. It should also be emphasised that bottles of wine are unstable in chemical terms and wine composition may change due to storage conditions. In contrast, the composition of stable isotopes is constant over time (they do not decay as a result of radioactive processes). Bearing in mind these considerations, methods based on stable isotope ratios are a logical choice in fraud detection applications. Indeed, since 1991 a method based on stable isotope ratios has been adopted as the official method in European Union regulations and by the Organisation Internationale de la Vigne et du Vin in order to detect fraud in the wine sector.

The climate of provenance areas affects the stable isotope ratios of wines, thus wines originating from different areas will be characterised by a different isotopic composition, affected by different climatic conditions. In addition to climatic and geographical dependence, the normal ranges of wine stable isotope ratios are different to the ratios

obtained in the case of wine samples produced with the addition of unpermitted substances during the winemaking process, such as cane sugar, beet sugar, water or mixtures of these. The isotopic parameters used for wine authentication are therefore selected in order to detect different kinds of common fraud in the wine sector, such as false claims of origin and the addition of unpermitted substances during the wine-making process.

In order to apply this method, a large databank is necessary. According to current regulations, at least 400 grape samples are collected every year in Italy. They must cover the full variability within all winegrowing regions, in terms of location, grape varieties, date of harvest, climatic variations and production technologies. To attest the authenticity of these samples, grapes are collected and not directly wine. Grapes are collected by officers of the Antifraud Department of the Italian Ministry of Food, Agriculture and Forestry and are then microvinified in controlled laboratory conditions, in compliance with the requirements of the EU reg 555/2008. The microvinification is performed without the addition of any unpermitted substances. Only *Saccharomyces cerevisiae* yeast is added to complete fermentation, following the official method. Once fermentation has been completed, the samples are used to measure the $(D/H)_1$, $(D/H)_2$, $\delta^{18}O$ and $\delta^{13}C$ parameters. The results obtained are used to compute normal ranges for authentic samples for all winegrowing regions separately. If the measured wine parameters for a wine sample under investigation fall outside these limits during wine authenticity assessments, the sample is not considered to be authentic.

5.2 Verification of claims of origin

The official method is based on stable isotope ratio analysis and univariate statistical analysis to detect fraud. In chapter 2, we carried out tests to see whether multivariate statistical methods are able to improve discrimination between wines produced in different Italian regions, helping to improve detection of false claims of origin. In this chapter the current practice of using five univariate t-tests for each isotopic parameter independently is extended and three more powerful statistical methods are provided.

During the procedure to check the labelled claim of origin for commercial wines, the $(D/H)_1$, $(D/H)_2$, $\delta^{18}O$ and $\delta^{13}C$ parameters are measured. These measured values are then compared with the normal ranges of samples from the wine databank, originating in the

same area that the wine under investigation claims to be from. If even one of the measured parameters of the wine under investigation is outside the normal range for that particular area, the wine is not considered to be authentic.

In our analysis, samples belonging to one administrative region were considered as a single class. Overall confidence intervals for all methods under evaluation were adjusted to 95% of probability, which is the percentage of samples belonging to a class correctly accepted as authentic. This is described as the sensitivity of the method. Since we set sensitivity to a constant value, specificity was calculated for each method and used for comparison. Specificity is the proportion of samples from other regions which are correctly detected as samples from outside the defined region. So higher specificity means better differentiation between samples from different regions and thus a higher probability of detecting false claims of origin.

The official method uses five univariate t-tests for each parameter separately. It is very important to set each separate t-test to 99% confidence, in order to achieve overall confidence of 95% ($0.99^5 = 0.95$).

Multivariate extension of t-tests involves the use of the Hotelling t-test, which uses the Mahalanobis distance. The Mahalanobis distance measures the distance of the tested sample from the centre of distribution of the corresponding class in multivariate space, taking into account correlations in the data. In this test the distribution of samples is assumed to be normal and very often samples are not normally distributed. Due to the difference in climatic and environmental factors, many samples from these heterogeneous regions have a multimodal distribution. As a result, confidence intervals are wider and statistical tests lose their power. To describe the multimodal distribution of these heterogeneous classes, a mixture modelling approach is applied. The mixture modelling approach uses several gaussians to model a class, the number being defined with criteria such as the Bayesian information criteria (BIC). The BIC strikes a balance between accurate description of the data and the number of parameters used. This approach is applied in both univariate and multivariate tests.

As expected, the results obtained confirmed that multivariate extension of univariate t-tests using the Mahalanobis distance have higher specificity and thus provide better discrimination between samples from different regions. When sample distribution is not normal, a mixture modelling approach obtained an additional improvement in univariate and multivariate approaches.

A particular advantage of the univariate and multivariate methods without mixture modelling is that no parameters need to be optimized, and for the mixture modelling we use the BIC criterion that is simple and fast. Many other methods require to perform a cross validation to determine the tuning parameters and this again would lead to a decrease in the number of samples on which the model is based. Very often in working practice there is limited number of samples per regions, therefore, the advantage of our proposed methods is that they can be applied directly in practice.

The performance of a multivariate method using the Mahalanobis distance for geographical differentiation of German wines was evaluated in the recent study by Monakhova et al. 2014, based on 718 wine samples. It showed good discrimination between neighbouring wine regions. Some overlaps between regions were also observed but this did not have a major implication in practice, due to the fact that the data are confidential and producers cannot know when the problem is present. Additional improvement in geographical differentiation was obtained with fusion of stable isotope data with ^{1}H NMR profiles.

Our study is recently confirmed (Jiang et al. 2015) on the basis of the study of Chinese wine samples, where a possibility for discrimination of wine samples originating from five different wine growing regions was investigated. Different subsets of stable isotope ratios ($(D/H)_1$, $(D/H)_2$, R, $\delta^{13}C$ and $\delta^{18}O$) were considered and perfect separation was achieved among wine samples from five regions using all five isotopic parameters. Results indicate that optimal separation can only be obtained when all five parameters are considered simultaneously with multivariate statistics.

5.3 Detecting the addition of sugar and water to wine

Addition of beet sugar, cane sugar or rectified must to grape must or wine before or during fermentation is carried out to increase the natural ethanol content in the wine. Dilution of wine with water is also used to increase the volume of wine and thus profit. These procedures are illegal and considered as a fraud in the wine sector. Because of changing stable isotope ratios values due to the addition of these substances, they are also used to detect this kind of fraud. In chapter 3, the possibilities of detecting the addition of sugar and water using stable isotope ratios based on a univariate method using five independent

t-tests and multivariate extension based on the Mahalanobis distance are evaluated. For this purpose, the addition of different amounts of cane sugar, beet sugar and water were simulated, in line with previous knowledge about this area and confirmed by experimental results.

Due to the difference in the $(D/H)_1$ ratio of ethanol from beet sugar and wines, the addition of beet sugar to grape must before fermentation significantly affects the parameter, in proportion to the amount of added beet sugar. The $(D/H)_2$ is not affected significantly because it mainly derives from the fermentation water. The $\delta^{13}C$ is not affected significantly due to the similarity of the $\delta^{13}C$ parameters of ethanol from beet sugar and wine, as both of them belong to the group of C_3 plants. The $(D/H)_1$ and $\delta^{13}C$ ratios of ethanol from cane sugar are significantly higher than those from wine, thus the addition of cane sugar affects both the $(D/H)_1$ and $\delta^{13}C$ ratios of pure wine, according to the proportion of added cane sugar. Dilution with water only causes a change to the $\delta^{18}O$ value.

Once again the results showed that the multivariate method is more effective in detecting adulteration. Cane sugar addition is easier to detect as compared to beet sugar, because two isotopic ratios change rather than one. For grape juices with low natural grape sugar content, due to climatic conditions such as lack of sunshine during grape development, even the addition of a small amount of sugar would lead to high adulteration fractions and thus detection of the fraud. This is the situation when some producers add sugar to the grape must. The results for simulations of water dilution show that it is easier to detect adulteration with water very distant from the normal range of water in wine, as expected.

For all applications, multivariate methods were more powerful in detecting non authentic samples than univariate statistical methods. Even in the case of multivariate methods some results may look too pessimistic, due to high variance within defined classes representing multivariate spaces where authentic samples should be. To decrease variance within classes, other climatic and geographical considerations should be borne in mind, as evaluated in chapter 4.

5.4 Climatic and geographical dependence of H, C and O stable isotope ratios of Italian wine

Due to strong climatic and thus geographic dependence, the stable isotope ratios of wine are adopted in many food traceability applications. There are many works confirming

strong relationships between the stable isotope ratios of wine and the climatic and geographical characteristics of provenance areas. In chapter 4, using a large databank with around 4000 samples collected over 11 different years, the isotopic ratios of wine (in particular $(D/H)_1$, $(D/H)_2$, $\delta^{13}C$ and $\delta^{18}O$) are evaluated in relation to a large number of climatic and geographical characteristics in the provenance areas. This is the largest study to date, in terms of the number of samples analysed and the climatic and geographical parameters used. Furthermore, the study was based on samples from a large, relative diverse climatic area, represented by Italy. Positive correlation between all wine isotopic parameters was observed. The isotope ratios had a tendency to decrease with an increase in latitude, distance from the sea and precipitation. Taking all the variables together, we come to other expected facts. All temperatures were highly correlated and when elevation increased temperatures decreased. Furthermore, harvesting periods for samples from areas at higher altitude were later in the year compared to samples from those at lower altitude. Using multiple linear regression we found that $\delta^{18}O$ was strongly linked to the provenance area ($R^2 = 0.71$) and was therefore shown to be the best individual marker for wine origin determination. $\delta^{18}O$ was followed by the $(D/H)_1$ parameter ($R^2 = 0.42$), $(D/H)_2$ ($R^2 = 0.30$) and $\delta^{13}C$ ($R^2 = 0.24$), which was the least related to wine provenance areas. Because of the nature of our study, in the field of correlation research, it was not possible to establish what really affects isotopic fractionation, different researchers providing different answers. More importantly, knowledge of those non-causal relationships is significant for sampling procedures and should be used as an additional consideration in fraud detection. Furthermore, knowledge of those relationships can be used to predict the isotopic composition of wines and to verify the declared origin of wines on the market, based on geographical coordinates and the climatic characteristics of the production regions.

5.5 Perspectives for future research

The performance of methods based on stable isotope ratios for the purposes of wine authentication is presented, with newly developed procedures for additional improvement. The results in chapters 2 and 3 suggest that even with the use of newly developed multivariate methods, a method based on five isotopic parameters has some limitations. In these situations there is sometimes an overlap between samples from different

geographical regions. This happens due to the similarity of climatic conditions in the two different regions, making it impossible to differentiate wine samples from these regions. To improve differentiation between wine samples from different regions, additional information variables such as elemental composition (Gremaud et all. 2004), ^{1}H NMR profiles (Monakhova et all. 2014) and strontium isotope ratios (Almeida and Vasconcelos 2001) should be tested together with the current parameters, taking into account the whole wine authentication context.

What is more, in some situations the detection level for a particular kind of adulteration (addition of beet sugar and water) is still high, allowing possible manipulation. To decrease the detection level in these cases, additional or new variables are indispensible.

However, it is necessary to pay attention. Additional variables might decrease the ability to detect the addition of sugar and water to wine, due to multiple testing corrections. The set of parameters should be selected with the scope of detecting at the same time false claims of origin and the addition of unpermitted substances (water, beet and cane sugar) during the winemaking process. With fewer parameters, statistical testing is more powerful.

Our statistical procedures provide improvement of the official method in the real context of wine authentication, which means that common frauds in wine sector (false claim of origin, sugar and water addition during winemaking process) can be detected more easily compared to the officially proposed procedure. Results of this work have been taken into consideration by the national experts in isotopic wine databank, and change of the European Regulation 555/2008 is necessary in order to use this approach in regulatory practice.

The stable isotope ratios of wine have been evaluated in relation to a large number of climatic and geographic parameters only partially evaluated in the past. Based on other literature in this field, several other research groups have confirmed the correlation between wine isotope ratios and many other parameters. Our study provides further insight into these phenomena, in the context of which climatic and geographic parameters are correlated to some extent, so we were unable to establish what affects wine isotopic fractionation. To obtain further insight into these phenomena, real experiments on a small scale should be performed.

In the event of missing data, these relationships may be used to predict the $\delta^{18}O$ composition of wines, using the appropriate climatic characteristics for the production region. This possibility should also be tested for other fresh agricultural commodities in the future.

References

Almeida, C.R.C. and Vasconcelos, M.T.S.D. (2001) ICP–MS determination of strontium isotope ratio in wine in order to be used as a fingerprint of its regional origin. Journal of Analytical Atomic Spectrometry 16, 607–611.

Gremaud, G., Quaile, S., Piantini, U., Pfammatter and E., Corvi, C. (2004) Characterization of Swiss vineyards using isotopic data in combination with trace elements and classical parameters. European Food Research and Technology 219, 97–104.

Jiang, W., Xue, J., Liu, X., Wang, D. I., Guo, Y. and Wang, L. (2015), The application of SNIF-NMR and IRMS combined with C, H and O isotopes for detecting the geographical origin of Chinese wines. International Journal of Food Science & Technology 50, 774–781.

Monakhova, Y.B., Godelmann, R., Hermann, A., Kuballa, T., Cannet, C., Schäfer, H., Spraul, M. and Rutledge, D.N. (2014) Synergistic effect of the simultaneous chemometric analysis of 1H NMR spectroscopic and stable isotope (SNIF-NMR, 18O, 13C) data: Application to wine analysis. Analytica Chimica Acta 833, 29–39.

Summary

The thesis deals with the H, C and O stable isotope ratios of wine and their application in wine authentication procedures. The aim of the thesis is to develop suitable statistical procedures able to improve current practice in official wine authentication procedures based on H, C and O stable isotope ratios. The second aim is to investigate to what extent the H, C and O stable isotope ratios of wine are related to the climatic and geographical characteristics of the provenance areas. No study considering such a large number of samples and variability factors and using such a complete statistical approach has been carried out to date.

In the introductory chapter 1, various methods used for wine authentication assessment are reviewed, with special attention for the method based on the stable isotope ratios of H, C and O officially adopted according to European Union regulations. An overview of current developments in the field of stable isotope ratios of bioelements and their application to wine authentication is also provided.

Chapter 2 evaluates univariate approaches underlying the official method to assess claims of origin, which is based on individual t-tests. To obtain better discrimination between samples from different regions, several extensions of univariate t-tests are proposed (one univariate and two multivariate statistical methods) and compared.

Chapter 3 evaluates performance in terms of the detection of unpermitted addition of cane and beet sugar to grapes before fermentation and dilution of wine with water using the univariate statistical method (based on t-tests) and also evaluates multivariate extension (Hotelling t-test). The addition of different amounts of water, cane and beet sugar to wine are simulated according to the experimental results, providing additional insight into the importance of different isotopic parameters for fraud detection.

Chapter 4 evaluates the relationships between wine $(D/H)_1$, $(D/H)_2$, $\delta^{13}C$ and $\delta^{18}O$ and the climatic and geographical parameters of provenance areas, as well as their general climatic dependence on wine growing areas. Comprehending these relationships is important for general understanding of isotopic phenomena, which can assist in making additional improvements to wine authentication procedures based on stable isotope ratios.

Finally, chapter 5 summarises all the findings resulting from this thesis. Perspectives for further research are also presented.

Samenvatting

In deze scriptie worden de stabiele H-, C- en O-isotoopverhoudingen behandeld evenals de toepassing ervan in wijnauthenticatieprocedures. In deze scriptie wordt ernaar gestreefd om geschikte statistische procedures te ontwikkelen die de huidige praktijken van de officiële wijnauthenticatieprocedures op basis van stabiele H-, C- en O-isotoopverhoudingen kunnen verbeteren. Bovendien wordt ernaar gestreefd om te onderzoeken in welke mate de stabiele H-, C- en O-isotoopverhoudingen van wijn gekoppeld zijn aan de klimatische en geografische eigenschappen van de herkomstgebieden. Tot nog toe is er geen enkel onderzoek uitgevoerd dat zo een groot aantal monsters en variabiliteitsfactoren beschouwt en dat zo een volledige statistische benadering gebruikt.

In het inleidende eerste hoofdstuk worden diverse methoden voor de authenticatiebeoordeling van wijn bekeken, met speciale aandacht voor de methode op basis van stabiele H-, C- en O-isotoopverhoudingen die officieel aangenomen is volgens de regelgeving van de Europese Unie. Er wordt ook een overzicht verstrekt van de huidige ontwikkelingen op het vlak van stabiele isotoopverhoudingen van bio-elementen. Ook wordt de toepassing ervan op de wijnauthenticatie verstrekt.

In hoofdstuk 2 worden univariate benaderingen geëvalueerd die aan de basis liggen van de officiële methode om de vermeende herkomst te beoordelen op basis van individuele t-tests. Om een betere onderscheiding tussen de monsters van verschillende regio's te verkrijgen worden er meerdere uitbreidingen van de univariate t-tests voorgesteld (één univariate en twee multivariate statistische methodes) en vergeleken.

In hoofdstuk 3 wordt de prestatie geëvalueerd op het vlak van de detectie van de verboden toevoeging van riet- en bietsuiker aan de druiven voor de gisting en de aanlenging van wijn met water. Dit gebeurt door middel van de univariate statistische methode (op basis van t-tests). Ook wordt de multivariate uitbreiding (t-test van Hotelling) geëvalueerd. De toevoeging van verschillende hoeveelheden water, riet- en bietsuiker aan wijn worden gesimuleerd volgens de resultaten van de experimenten en verstrekken zo extra inzicht in het belang van verschillende isotopische parameters voor de opsporing van fraude.

Hoofdstuk 4 evalueert de relaties tussen wijn $(D/H)_1$, $(D/H)_2$, $\delta^{13}C$ en $\delta^{18}O$ en de klimatische en geografische parameters van de herkomstgebieden, evenals hun algemene klimatische afhankelijkheid van wijnbouwgebieden. Het begrijpen van deze relaties is

belangrijk voor het algemeen begrip van isotopische fenomenen, wat kan bijdragen tot het doorvoeren van bijkomende verbeteringen aan de wijnauthenticatieprocedures op basis van stabiele isotoopverhoudingen.

Ten slotte worden in hoofdstuk 5 alle bevindingen van deze scriptie samengevat. Er worden ook perspectieven voor verder onderzoek voorgesteld.

Curriculum vitae

Nikola Dordevic was born on 17 August 1980 in Smederevo, Serbia. He finished his elementary and secondary education in Smederevo. Following this, he graduated from the department of Food Technology in the Faculty of Agriculture at University of Belgrade in 2006. He obtained his M.Sc. degree in Food Science on 2010, from the master program jointly organized by the Catholic University of Leuven (KU Leuven) and the Ghent University (UGent), Belgium. He started Ph.D. studies on 2011 and it was a joint collaboration between Fondazione Edmund Mach (Trento, Italy) and the Analytical Chemistry/Chemometrics group from Radboud University Nijmegen (the Netherlands) where he worked on the application of chemometrics in the area of stable isotope ratios of bioelements. Nikola started his collaboration with Centre for Computational and Systems Biology (COSBI) in November 2014.

Acknowledgments

First I would like to express my gratitude to my supervisors Federica Camin and Ron Wehrens for their guidance and assistance in this research. I thank them for their confidence in me throughout the PhD programme.

I would also like to thank my promoter Lutgarde Buydens for her constructive and positive instructions and suggestions and Geert Postma for his kind support, valuable advice and help.

It would have been impossible to successfully finish this thesis without many other people from the Fondazione Edmund Mach. I thank my colleagues in the Biostatistics and Data Management Research Group, Matthias Scholz, Pietro Franceschi, Nir Shahaf, Mridula Prasad, Marco Giordan, Dong Yonghui and Nay Min Min Thaw Saw, for their invaluable support.

My thanks go to those involved in the Stable Isotope and Traceability Platform, Luana Bontempo, Marco Simoni, Agostino Tonon, Katryna Vanleeuwen, Luca Ziller and Mateo Perini, for their kindness, friendship and support during my PhD.

My sincere thanks also go to Markus Neteler and Luca Delucchi from the GIS and Remote Sensing platform for useful discussions and support with climatic data.

I would also like to express my appreciation to all those in the Analytical Chemistry/Chemometrics Group, Jasper Engel, Jeroen Jansen, Lionel Blanchet, Ewa Szymanska, Brigitte Geurts, Agnieszka Smolinska, Brigitte Loozen and Jan Gerretzen, for their understanding and for creating a friendly environment during my stay at Radboud University.

I thank all my colleagues and friends in the “FEMous” basketball team, especially Panagiotis Arapitsas, Tomas Roman, Chidananda Nagamangala, Andrea Cattani and Alberto Maria Cattaneo. It was a pleasure to study and work with you.

I thank the administrative staff of the PhD and welcome offices of the Fondazione Edmund Mach, especially Alessandro Gretter and Elisabetta Perini, for their kind support during my PhD studies.

I gratefully acknowledge the financial support of the Grana Padano Consortium for funding this study.

I also wish to express my deepest gratitude to my family, my father Milos and my mother Branka for supporting me spiritually throughout my life. Last but not the least, the most important thanks go to my girlfriend Vesna. Thank you for everything.